# Charles University

Faculty of Social Sciences
Institute of Economic Studies

MASTER'S THESIS

# The Elasticity of Substitution between Skilled and Unskilled Labor: A Meta-Analysis

Author: **Bc. Ľubica Laslopová**

Supervisor: **PhDr. Zuzana Havránková, Ph.D.**

Academic Year: **2017/2018**

## Declaration of Authorship

The author hereby declares that she compiled this thesis independently, using only the listed resources and literature, and the thesis has not been used to obtain a different or the same degree.

The author grants to Charles University permission to reproduce and to distribute copies of this thesis document in whole or in part.

Prague, January 3, 2018

Signature

## Acknowledgments

I am especially grateful to PhDr. Zuzana Havránková, Ph.D. and doc. PhDr. Tomáš Havránek, Ph.D. for their useful advice that really made it all seem easier. Also, I owe my gratitude to my family and friends for supporting me through my studies.

## Abstract

In this thesis we use meta-analytic methods to quantitatively summarize empirical evidence on elasticity of substitution between skilled and unskilled labor. Review is based on sample of 684 estimates from 78 studies. After a brief overview of theoretical framework, estimation strategies and distribution of the existing estimates, we test for publication bias. According to regression-based tests for publication bias, we do not reject null hypothesis of no publication bias in the existing literature. To explain heterogeneity between the estimates, we use Bayesian Model Averaging. We find that both real factors and research design influence resulting value of estimated elasticity. Our synthetic estimates of the elasticity imply that skilled and unskilled workers are imperfect substitutes in the long run, substitutability in the short-run is mostly limited.

| | |
|---|---|
| **JEL Classification** | J82, J23, J24, J31 |
| **Keywords** | elasticity of substitution between skilled and unskilled labor, meta-analysis, publication bias |

| | |
|---|---|
| **Author's e-mail** | `llaslopova@gmail.com` |
| **Supervisor's e-mail** | `zuzana.havrankova@fsv.cuni.cz` |

## Abstrakt

V tejto práci sú použité metódy meta-analýzy za účelom kvantitatívneho výskumu empirickej literatúry, ktorá sa zaoberá elasticitou substitúcie medzi kvalifikovanou a nekvalifikovanou prácou. Analýza je založená na 684 odhadoch zo 78 štúdií. Po stručnom predstavení teórie, rozličných stratégií odhadu a distribúcie existujúcich odhadov, testujeme existenciu systematickej odchýlky ako následku publikačnej selektivity. S použitím testov založených na regresnej analýze sa nepodarilo zamietnuť nulovú hypotézu neexistencie systematickej odchýlky následkom publikačnej selektivity. Za účelom skúmania rozdielov medzi odhadmi je použité Bayesovské priemerovanie modelov. Použitím tejto metódy zisťujeme, že odhady sú ovplyvnené nielen faktormi, ktoré majú dopad na skutočnú hodnotu elasticity, ale aj metódami výskumu a použitými dátami. Naše syntetické odhady elasticity naznačujú, že kvalifikovaní a nekvalifikovaní pracovníci sú v dlhodobom časovom horizonte nedokonalými substitútmi. V krátkodobom časovom horizonte je ich substitúcia limitovaná.

**Klasifikace JEL** J82, J23, J24, J31
**Klíčová slova** elasticita substitúcie medzi kvalifikovanou a nekvalifikovanou prácou, metaanalýza, publikačná selektivita

**E-mail autora** llaslopova@gmail.com
**E-mail vedoucího práce** zuzana.havrankova@fsv.cuni.cz

# Contents

# List of Tables

# List of Figures

# Acronyms

**2SLS** Two-stage Least Squares

**AES** Allen-Uzawa Elasticity of Substitution

**BMA** Bayesian Model Averaging

**CES** Constant Elasticity of Substitution

**FAT** Funnel Asymmetry Test

**FE** Fixed Effects

**HES** Hicks Elasticity of Substitution

**ME** (Multilevel) Mixed Effects

**MCMC** Markov Chain Monte Carlo

**OLS** Ordinary Least Squares

**PET** Precision Effect Test

**PIP** Posterior Inclusion Probability

**PMP** Posterior Model Probability

# Master's Thesis Proposal

| | |
|---|---|
| **Author** | Bc. Ľubica Laslopová |
| **Supervisor** | PhDr. Zuzana Havránková, Ph.D. |
| **Proposed topic** | The Elasticity of Substitution between Skilled and Unskilled Labor: A Meta-Analysis |

**Motivation** Since the emergence of labor economics, considerable progress has been made concerning role of education in the economy. Some concepts, such as wage returns to education or increase in labor productivity due to additional human capital, rely on assumptions about substitutability of workers with different levels of education.(Freeman, 1986), (Behar, 2010) Therefore there exist numerous scientific studies dealing with the topic of qualified workers replacing the unqualified and vice versa. One approach to study this phenomenon is to quantify it by measuring elasticity of substitution between the two groups. In my bachelor thesis, I focused on overeducated individuals, that is, workers whose education exceeds the level required for their job. The increasing evidence of overeducation and undereducation suggests that the elasticity of substitution might be higher that it has been assumed earlier. So what can we learn from the latest studies of the elasticity? Do labor economists need an update of assumptions about the elasticity of substitution between skilled and unskilled labour? This thesis aims to provide an answer to these questions through a systematic investigation of relevant studies.

## Hypotheses

1. Hypothesis #1: The elasticity of substitution between more and less qualified workers is in the range of $0.5 - 2$ as suggested by earlier studies.
2. Hypothesis #2: The estimate of elasticity significantly depends on estimation method used by the study.
3. Hypothesis #3: The estimate of elasticity significantly depends on characteristics of the country studied.

4. Hypothesis #4: Substitutability of skilled and unskilled labor is higher in the long run than in the short-run.

5. Hypothesis #5: Accounting for the capital-skill complementarity effect has an impact on estimate of elasticity.

6. Hypothesis #6: There is no publication bias - the elasticity estimates are not systematically shifted due to selective reporting.

**Methodology** Firstly we will accumulate results of multiple studies and different characteristics of these studies (such as used methods, data types, control variables or countries studied. Next we will construct our dataset. To assess which characteristics of the research have impact on its result we will use Bayesian Model Averaging (BMA), described for instance by Hoeting et al. (1999). This approach prevents from underestimating uncertainty in the estimated coefficients, as the posterior distribution of the estimated quantity is a weighted average of the posterior probability distributions under each possible model given a set of potential predictors. (Raftery et al., 1997) Our own estimation of the elasticity based on summarized evidence will follow. Rather than simply taking mean of all estimated elasticities of substitution, we will try to define "best practice" in literature and generate our estimate as a fitted value using maximum value for every aspect of the study which is preferred and minima for the unwanted aspects. Publication bias will be tested following the example of Egger et al. (1997) by examining the assymetry of the so-called funnel plot and, more formally, statistically examining the relationship between estimates and standard errors.

**Expected Contribution** The estimated elasticity of substitution between skilled and unskilled labour varies from one study to another. Therefore a meta-analysis is needed to summarize and compare these findings in a structured and objective way and eventually correct for systematic mismeasurements.This helps readers to better understand differences between results and correctly interpret existing evidence .

## Outline

1. Introduction

2. Context and Importance of Elasticity Estimation

3. Methods of the Elasticity estimation

4. Dataset collection and description

5. Publication bias and corrective measures

6. Results

7. Conclusion

## Core bibliography

Autor, David H. ,"Skill Biased Technical Change and Rising Inequality: What is the Evidence? What are the Alternatives?" MIT and NBER (2002)

Behar, Alberto."The elasticity of substitution between skilled and unskilled labor in developing countries is about 2."; (2010).

Ciccone, Antonio, and Giovanni Peri. "Long-run substitutability between more and less educated workers: evidence from US states, 1950 −1990." Review of Economics and statistics 87.4(2005) : 652 − 663.

Egger, Matthias, et al."Bias in meta-analysis detected by a simple, graphical test." Bmj 315.7109 (1997): 629 − 634.

Freeman, Richard B. "Demand for education." Handbook of labor economics1. (1986): 357 − 386.

Hoeting, Jennifer A., et al. "Bayesian model averaging: a tutorial." Statistical science (1999): 382 − 401.

Katz, Lawrence F., and Claudia Dale Goldin. "The race between education and technology." Cambridge: Harvard University Press, 2008.

Raftery, Adrian E., David Madigan, and Jennifer A. Hoeting. "Bayesian model averaging for linear regression models." Journal of the American Statistical Association 92.437 (1997): 179 − 191.

Author

Supervisor

# Chapter 1

# Introduction

Results from empirical studies in economics often seem to be contradictory. Though they aim to estimate the same quantity or effect, results can vary greatly. Parameters of the production function are not an exception. No individual study can provide definitive answer to complex economic questions linked with the interpretation of these parameters. Elasticity of substitution between skilled and unskilled labor is an important parameter widely used to model various labor market settings. First, substitutability between differently skilled workers plays a crucial role in the branch of research exploring wage inequality. The extent to which wage differential reacts to the change in relative supply of skilled labor directly depends on the value of elasticity of substitution (Behar 2009). Policymakers may be interested in addressing wage inequality by creating incentives to raise overall education level. Substitutability between skill categories can provide some information about whether these attempts are likely to have intended effects.

Another application of the elasticity is linked with capital-skill complementarity hypothesis. The idea is that technological progress may have different effect on groups with different skills and knowledge. If a shift in production technology increases the relative productivity of skilled labor and consequently leads to an increase in relative demand, this technological change is said to be skill-biased. Greater increase in productivity of highly skilled group is causing wages of highly skilled workers to grow faster. Inequality between the two groups is increased if the effect of skill-biased technical change is not offset by another economic force with opposite direction - increase in overall education level. What is the role of the elasticity in this process? *Ceteris paribus*, the higher the elasticity of substitution between skilled and unskilled labor, the stronger will be the increase in inequality between differently skilled groups attributable to skill-biased technical change. Also, the effect of the increase in overall education level is weaker if substitutability is high (Kierzenkowski & Koske 2012). Numerous empirical studies provide evidence for the existence of skill-biased technological change, for example (Krusell *et al.* 2000) or (Katz & Murphy 1992).

Researchers attempted to estimate the elasticity of substitution between skill groups since 1970. In *Demand for Education*, Freeman (1986) provided a brief overview of 11 studies that reported values ranging from values close to 0 to as much as 1000. Hamermesh (1996) collected about 40 studies concerning skill heterogeneity between workers and provided estimates ranging from approximately -0.5 to 6.0. However, these overviews of the literature lack further analysis aiming to cope with possible publication bias, differences between studies and to answer the fundamental question: are skilled workers substitutes or complements? And if they are substitutable, to what extent?

The objective of this thesis is to provide systematic quantitative overview of existing literature dealing with the estimation of elasticity of substitution between skilled and unskilled labor. For this purpose, we apply methods of meta-analysis. Meta-analysis is a synthesis of the results of separate but comparable scientific studies. It applies statistical methods to explain differences between these studies, to identify potential systematic bias and eventually to derive synthetic estimates. The term meta-analysis was first used in 1978 by Gene V. Glass, who collected and analyzed results of 375 studies concerning psychotherapy. Since then, meta-analysis has been used in medical research and also in social sciences including economics.

Although there are articles that summarize existing estimates, there is not yet any meta-analysis dealing with the empirical literature about substitution elasticity between skilled and unskilled labor. To our best knowledge, we are the first to test for publication selection bias and to inspect how characteristics of studies and estimates themselves relate to magnitude of the estimated elasticity. For this purpose, we apply method called Bayesian Model Averaging which we believe to be superior to fitting a single model, because it explicitly accounts for model uncertainty.

The thesis is structured the following way: In Chapter 2 we discuss theoretical framework behind the elasticity of substitution between skilled and unskilled labor. We present definition of the elasticity and its different measures. Then we focus on different forms of the production function used in the literature.

In Chapter 3, we follow with a brief overview of empirical estimation where we summarize different estimation strategies and possible issues of the estimation. Chapter 4 describes how the estimated elasticities, their standard errors and additional characteristics of studies and estimates were collected, presents descriptive statistics and provides additional information about adjustments to the constructed dataset. Chapter 5 deals with the possibility of publication selectivity bias in existing literature. After a definition of publication selection bias and overview of methodology behind its diagnostics, we present funnel plot of existing estimates and conduct formal, regression-based tests for publication bias. In Chapter 6 we aim to explain what drives heterogeneity between the existing estimates. Finally, Conclusion provides a brief overview of the results and concluding remarks.

# Chapter 2

# Theoretical Framework

## 2.1 Definition of Elasticity of Substitution Between Skilled and Unskilled Labor

Concept of the elasticity of substitution was first introduced by Hicks (1932) who studied substitutability between capital and labor in production. Since then, multiple definitions of substitution elasticity emerged. Common aim is to measure how easily a factor of production can be replaced for other factors, holding the output fixed. But there are differences including treatment of different variables as exogenous (either prices or quantities), whether they measure absolute or relative changes, and whether the other factors are allowed to vary or not. These differences are closely linked to choices of estimation strategies, as we will see later in this chapter. We will only discuss measures used in reviewed literature, an extensive review is provided for example by Broadstock *et al.* (2007) or Zachłod-Jelec & Boratyński (2016).

Most researchers estimating the elasticity of substitution between skilled and unskilled labor define it as percentage change of ratio in which these two factors are used divided by percentage change of the ratio of their marginal products. For instance, this relationship is estimated by Katz & Murphy (1992), Ciccone & Peri (2005) or Razzak & Timmins (2008). This definition of the elasticity is frequently referred to as "Hicks elasticity of substitution" (HES). We use this notation, though in fact, this relationship has been defined by Robinson (1969) and its equivalence with the original definition by Hicks is only true in the two-factor case.

Mathematically written, if $x_1$ and $x_2$ are two factors of production, $f()$ is a strictly quasi-concave production function homogeneous of degree $k > 0$[1], $f_1$and $f_2$ are the first derivatives of the function $f()$ with respect to $x_1$ and $x_2$, respectively, and $\sigma_{HES}$ is the elasticity of substitution between them, then the elasticity of substitution is defined as follows:

$$\sigma_{HES,21} = \frac{\frac{d(x_2/x_1)}{x_2/x_1}}{\frac{d(f_1/f_2)}{(f_1/f_2)}} \tag{2.1}$$

Output is held constant, quantities are assumed to be endogenous, prices exogenous. It is asymmetric, meaning that $\sigma_{HES,21}$ is not necessarily equal to $\sigma_{HES,12}$. This definition of the elasticity of substitution between two factors allows to measure curvature of isoquants of the production function.[2] The more curved the isoquant is, the lower the elasticity ($\sigma_{HES}$) (Broadstock *et al.* 2007). As the production function is strictly quasi-concave, elasticity of substitution can possibly be any real number from the interval $(0, \infty)$, the two extreme cases are: L-shaped (or Leontief) isoquants corresponding to zero elasticity and linear, constantly downward sloping isoquants in case of infinite elasticity.

Assuming perfect factor markets, production factors are paid exactly their marginal products. Then the elasticity also measures percentage change of the ratio of product quantities to the percentage change of the ratio of their respective prices. In case that $w_1$ and $w_2$ stand for wage levels of skilled and unskilled labor, respectively, and $MP_1$ and $MP_2$ for their marginal products, it can be written:

$$\sigma_{HES,21} = \frac{\frac{d(x_2/x_1)}{x_2/x_1}}{\frac{d(MP_1/MP_2)}{MP_1/MP_2}} = \frac{\frac{d(x_2/x_1)}{x_2/x_1}}{\frac{d(w_1/w_2)}{w_1/w_2}} = \frac{d\ ln(x_2/x_1)}{d\ ln(w_2/w_1)} \tag{2.2}$$

The logarithm form of the equation is frequently used in empirical studies.

What is the economic interpretation of this relationship? In case that $\sigma = 0$, the two kinds of labor are perfect complements. Fixed proportions of the two are needed to increase production level and they cannot be substituted for each other. If the elasticity of substitution is in the interval $(0, 1)$, differently skilled workers are complements, their contribution to resulting output is different. Elasticity equal to 1 is called unitary elasticity and translates to relative quan-

[1]Eventually, elasticity of substitution can be defined for any concave production function with two or more variables.

[2]Isoquant being a contour line of the production function that defines set of combination of the inputs producing the same amount of output.

tity change to be exactly proportional to relative price change. If the elasticity of substitution is in the interval $(1, \infty)$, skilled and unskilled workers are substitutes. Unskilled workers can easily work on positions intended for skilled workers and skilled workers, on the contrary, may easily be overeducated for their position. Infinite elasticity means that the two kinds of labor are perfect substitutes. In this case, skill is no longer an important factor for the resulting output and skilled and unskilled labor are economically the same commodity. In theory, this implies that in case that wage of either type of labor is higher, this type of labor will not be used in production. On the contrary, in case of near zero elasticity, the two types of labor are practically impossible to substitute, fixed proportions of the two are needed to effectively produce the output.

An alternative measure of the substitution elasticity is called Allen-Uzawa Elasticity of Substitution. Instead of derivatives of the production function, it is defined using cost function. If cost function is denoted by $c$, its partial derivatives and cross-partial second derivative as $c_1$, $c_2$ and $c_{12}$, respectively, then we can write:

$$\sigma_{AES,21} = \frac{cc_{12}}{c_1 c_2} \tag{2.3}$$

As suggested by Gyimah-Brempong & Gyapong (1992), it can also be written in terms of cross price elasticity scaled by cost share:

$$\sigma_{AES,21} = \frac{d\ lnx_1 / d\ lnw_2}{s_2} \tag{2.4}$$

Important difference from previously defined Hicks elasticity of substitution is that values of Allen-Uzawa elasticity can be negative. If $\sigma_{AES} > 0$, factors are classified as substitutes and if $\sigma_{AES} < 0$, factors are classified as complements. Cost share being nonnegative, $\sigma_{AES}$ has always the same sign as cross price elasticity. Consequently, the interpretation is different. It does not measure changes of relative quantities moving along an isoquant of production. Instead, as we can see from the second form, it is a one-factor-one-price elasticity. Unlike $\sigma_{HES}$, elasticity is symmetrical. Allen-Uzawa Elasticity of substitution is widely used in empirical studies because it can be relatively easily derived from estimated cost function (Broadstock *et al.* 2007). This type of elasticity of substitution between skilled and unskilled labor is estimated for example by Bergström & Panas (1992), Askilden & Nilsen (2005) or Jensen & Morrisey (1986).

Few authors use alternative measures of elasticity other than Hicks elasticity or Allen-Uzawa elasticity. Apart from $\sigma_{AES}$, Gyimah-Brempong & Gyapong (1992) estimate so-called Morishima elasticity of substitution:

$$\sigma_{MES,21} = \frac{d\ ln(\frac{x_2}{x_1})}{d\ lnw_2} \tag{2.5}$$

Morishima elasticity of substitution is similar to $\sigma_{HES}$: is also asymmetrical and it cannot be negative. But the difference is that it is a two-factor-one-price elasticity, so the measure tracks change in factor ratio in response to change in one price of one factor.

Finally, Berndt & Christensen (1974) and Gyimah-Brempong & Gyapong (1992) estimate so-called Shadow elasticity of substitution. It is in fact a special case of $\sigma_{HES}$, but allows prices to change such that average cost is fixed (Gyimah-Brempong & Gyapong 1992).

## 2.2 Production Functions

For a long time, production functions implicitly treated labor as homogeneous. Differences in education, skill and experience were ignored. To begin with, in the well-known Cobb-Douglas case[3], skilled and unskilled labor are both aggregated into one production factor assuming infinite elasticity. Elasticity of substitution between labor and capital, on the other hand, is equal to one.

Later on, economists started to discuss the role of education and skill in individual decision-making and aggregate production process. The concept of human capital, introduced by Gary S. Becker in 1964 enhanced branch of research dedicated to the impact of education and skill on production process. Accumulation of knowledge and skill is assumed to increase marginal product and consequently, wage of a worker. Important question emerged: What is a contribution of skill to national output? An extension of production function was needed to answer the question. One step to answer the question is to divide workforce into two or more heterogeneous subgroups based on the level of skill. Then their combination can be modeled by multiple frameworks. In the reviewed literature, authors typically assume one of the following forms:

[3] Defined as $Y = c \cdot L^{\alpha} \cdot K^{\beta}$, where $L, K$ stand for labor and capital, respectively, $\alpha$, $\beta \in [0, 1]$ and $c$ is a constant different from zero.

- One-level CES function
- Multilevel (or nested) CES function
- Translog production function
- Translog cost function

### 2.2.1 One-level CES Function

In general, Constant elasticity of substitution (CES) functional form is defined as:

$$Y_t = [\alpha_t(a_t X_{it})^\rho + (1-\alpha_t)(b_t X_{jt})^\rho]^{\frac{1}{\rho}} \tag{2.6}$$

where $X_{it}$ and $X_{jt}$ are two factors of production, $a_t$ and $b_t$ are factor augmenting technology indexes and $\alpha_t$ is a technology parameter interpretable as indexing the "share of work" allocated to the factor $X_i$, at time $t$. This two-factor CES production function was introduced by Solow (1956) and it originally included capital and labor ignoring skill differences between workers.

For the sake of simplicity, some authors only take the basic CES form (3.1) and instead of labor and capital, they plug in skilled employment ($L_S$) and unskilled employment ($L_U$) as production factors. This way, skilled and unskilled labor are treated as the only factors producing the output. Then the starting point for estimation of elasticity is:

$$Y_t = [\alpha_t(a_t L_{St})^\rho + (1-\alpha_t)(b_t L_{Ut})^\rho]^{\frac{1}{\rho}} \tag{2.7}$$

An important feature of CES functional form is that the elasticity is constant along the relative demand curve (Ciccone & Peri 2005). In other words, elasticity of substitution is constant irrespective of changes in relative supply. This elasticity can be easily derived from the parameter $\rho$ as

$$\sigma = \frac{1}{1-\rho} \tag{2.8}$$

### 2.2.2 Multilevel CES Function

Another class of production functions used in literature also assumes CES technology, but applies a nested structure. There are multiple ways to nest production factors under CES framework. An important feature of nested

CES functional form is that choice of the nesting structure impose additional restriction on elasticities. In case of three factors, researchers estimate two elasticities, the third one is given by the symmetry of CES structure (Krusell *et al.* 2000). For instance, if $CES_1$ and $CES_2$ are two functions of CES type and $x_1$, $x_2$ and $x_3$ are three production factors, nesting structure of the form $Y = CES_1(x_1,\ CES_1(x_2, x_3))$ would imply that the elasticity of substitution between $x_1$ and $x_2$ is equal to the elasticity of substitution between $x_1$ and $x_3$. Therefore the choice of nesting structure is not arbitrary.

Usually, three production inputs are considered: skilled labor, unskilled labor and capital. One stream of literature assumes output to be a function of capital and labor at the first level and decomposes labor to skilled and unskilled at the second level. For instance, Avalos & Savvides (2006) assumes production function in the following form:

$$Y_t = K_t^{\beta}[\alpha_t(a_t L_{St})^{\rho} + (1 - \alpha_t)(b_t L_U)^{\rho}]^{\frac{1-\beta}{\rho}} \tag{2.9}$$

where $K$ stands for capital and $\beta$ is the share of capital in production. In this case, the first level that models factor contribution of skilled and unskilled labor is assuming CES technology. The second level, or "aggregator" is in Cobb-Douglas form. This restricts the elasticity between labor and capital to be equal to one. As in case of one-level CES, the elasticity of substitution is defined by: $\sigma = \frac{1}{1-\rho}$ .

Borjas & Katz (2007) or Borjas (2003) also assume three factor case with capital, skilled and unskilled labor. But the second-level aggregator is of type CES:

$$Y_t = [\delta_t(K_t)^{\beta} + (1 - \delta_t)(L_t)^{\beta}]^{\frac{1}{\beta}}, \tag{2.10}$$

where $L_t$ stands for aggregate labor and is defined as:

$$L_t = [\alpha_t(a_t S_t)^{\rho} + (1 - \alpha_t)(b_t U_t)^{\rho}]^{\frac{1}{\rho}} \tag{2.11}$$

Using this functional form, elasticity of substitution between labor and capital is no longer restricted to be equal to one, but this form does not allow for capital-skill complementarity as it implicitly assumes that capital is equally substitutable for skilled and unskilled labor.

Other studies apply an alternative nesting scheme where workers are divided into more than two levels according to their skill. For instance, in study by Manacorda *et al.* (2010), the first level relates output and two kinds of labor

- skilled and unskilled. The second level then decomposes skilled labor to secondary educated and tertiary educated labor:

$$Y_t = A_t(\alpha_t \cdot L_{st}^{\rho} + (1 - \alpha_{Ht}) \cdot L_{ut}^{\rho})^{1/\rho}, \tag{2.12}$$

and

$$L_{st} = B_t(\alpha_{2t} \cdot L_{2t}^{\gamma} + (1 - \alpha_{2t}) \cdot L_{3t}^{\gamma})^{1/\gamma}, \tag{2.13}$$

where $\alpha_{2t}$ stands for relative productivity of skilled workers at time $t$ and $\alpha_{2t}$ stands for relative productivity of secondary educated workers at time $t$. Parameter $A_t$ may be interpreted as skill-neutral technological change.

Perhaps the most complicated nesting structure is used by Krusell *et al.* (2000) whose example has been followed by Lindquist (2004) and Dupuy (2007). They assumed four production factors - capital structure, capital equipment, skilled labor and unskilled labor and three-level nesting structure.

### 2.2.3 Translog framework

Another type of production function used in reviewed literature is called translog production function. Translog function is an abbreviation for transcendental logarithmic function and it is the second-order approximation of CES function by Taylor polynomial about $\rho = 0$ (Berndt & Christensen 1973). Unlike CES framework, when translog form is used, elasticity of substitution is allowed to vary with relative supply of differently educated workers (Ciccone & Peri 2005). Assuming three production factors: capital ($K$), skilled and unskilled labor ($L_S$) and ($L_U$), respectively, translog production function can be written as follows:

$$\begin{aligned} ln(Y) = ln\alpha_0 + \alpha_s lnL_S + \alpha_u lnL_U + \alpha_k lnK + \frac{1}{2}\gamma_{uu}(lnL_U)^2 + \gamma_{us} lnL_U lnL_S \\ + \gamma_{uk} lnL_U lnK + \frac{1}{2}\gamma_{ss}(lnL_S)^2 + \gamma_{sk} lnL_S lnK + \frac{1}{2}\gamma_{kk}(lnK)^2 \end{aligned} \tag{2.14}$$

Duality of profit maximization and cost minimization framework allows the use of translog cost function instead of translog production function:

$$
\begin{aligned}
ln(C) = ln\alpha_0 + \alpha_s lnw_S + \alpha_u lnw_U + \alpha_k lnr + \alpha_y lnY + \frac{1}{2}\gamma_{uu} lnw_U^2 + \frac{1}{2}\gamma_{us} lnw_U lnw_S \\
+\frac{1}{2}\gamma_{uk} lnr lnw_S + \frac{1}{2}\gamma_{ss}(lnw_S^2) + \frac{1}{2}\gamma_{sk}(lnw_S lnr) + \frac{1}{2}\gamma_{kk}(lnr)^2 \\
+\gamma_{sy} lnw_S lnY + \gamma_{uy} lnw_U lnY + \gamma_{ky} lnr lnY + \frac{1}{2}\gamma_{yy} lnY lnY
\end{aligned}
\tag{2.15}
$$

Here, $C$ denotes total costs, $w_S$,$w_U$ and $r$ stand for wages of skilled labor, unskilled labor and cost of capital (rent). An important intermediate step in elasticity estimation is to derive cost share equations. Using Sheppard's lemma[4], it can be written:

$$s_s = \alpha_s + \gamma_{sy} lnY + \gamma_{su} lnw_U + \gamma_{sk} lnr, \tag{2.16}$$

$$s_u = \alpha_u + \gamma_{uy} lnY + \gamma_{us} lnw_S + \gamma_{uk} lnr, \tag{2.17}$$

and

$$s_k = \alpha_k + \gamma_{ky} lnY + \gamma_{ks} lnw_S + \gamma_{ku} lnw_U \tag{2.18}$$

where $s_s$, $s_u$ and $s_k$ are cost shares of skilled, unskilled labor and capital, respectively (Bergström & Panas 1992). Now, there are more possible measures of elasticity that can be derived using cost share equations. In terms of the above mentioned relations, Allen-Uzawa elasticity of substitution of skilled and unskilled labor is:

$$\sigma_{AES,su} = 1 + \frac{\gamma_{us}}{s_s s_u} \tag{2.19}$$

Given the cost-share equations, Morishima elasticity can be also easily derived as follows:

$$\sigma_{MES,su} = \frac{\alpha_{su}}{s_s} + s_u - (\frac{\alpha_{us}}{s_u} + s_s) \tag{2.20}$$

Alternatively, assuming translog cost function, Shadow elasticity can be estimated. Gyimah-Brempong & Gyapong (1992) derive it by plugging in Allen-Uzawa elasticity into the following formula:

$$\sigma_{SES,su} = \frac{S_s S_u}{S_s + S_u}(2\sigma_{AES,su} - \sigma_{AES,ss} - \sigma_{AES,uu}) \tag{2.21}$$

[4]According to Sheppard's lemma, conditional factor demand equals partial derivative of the cost function with respect to price of this particular factor.

Most researchers working with translog framework use cost function instead of translog production function. It is probably for its convenience: with translog cost function, elasticity estimation is more straightforward. However, there are exceptions. For instance, Jensen & Morrisey (1986) assume translog production function and derive the elasticity using marginal products according to McFadden's formula for partial elasticity of substitution.

# Chapter 3

# Estimation of the Elasticity of Substitution

## 3.1 From Production Functions to Estimated Equations

Most researchers assume production function of type CES. Whether it is one-level CES function or CES with two and more levels, the estimation procedure is mostly similar[1]. For the sake of simplicity, we will derive estimation equation for one-level CES. The following steps are needed to obtain commonly estimated equation. First, marginal products are obtained by taking derivatives of the equation (2.7) with respect to $S_t$ and $U_t$. Assumption of competitive labor markets allows to write the equality between wage ratio and ratio of marginal products:

$$\frac{w_S}{w_U} = \frac{\alpha_t \; a_t^{\rho} S_t^{\rho-1}}{(1-\alpha_t) b_t^{\rho} U_t^{\rho-1}} \tag{3.1}$$

Substitution for $\rho$ using (2.8) and slight rearrangement leads to:

$$\frac{w_S}{w_U} = \frac{\alpha_t}{(1-\alpha_t)} (\frac{a_t}{b_t})^{\frac{\sigma-1}{\sigma}} (\frac{S_t}{U_t})^{-\frac{1}{\sigma}} \tag{3.2}$$

Taking logarithms of (3.2) leads to the final specification:

[1] Simple estimation equations similar to one-level CES case can be found, for instance, in Katz & Murphy (1992), Gallego (2012) or (Borjas 2003). An exception is the study of (Krusell *et al.* 2000), who use structural equations to estimate parameters of their model that was primarily designed to explore capital-skill complementarity.

$$ln(\frac{w_S}{w_U}) = \beta_0 + \beta_1 ln(\frac{a_t}{b_t}) + \beta_2 ln(\frac{S_t}{U_t}) \tag{3.3}$$

with $\beta_0 = ln(\frac{\alpha}{1-\alpha})$, $\beta_1 = \frac{\sigma-1}{\sigma}$ and $\beta_2 = -\frac{1}{\sigma}$. Coefficient of interest is $\beta_2$, and may be interpreted as the estimated effect of relative supply of skilled labor on wage premium to skilled workers. Researchers often include other regressors to capture different characteristics of workers or labor markets, leading to the following equation:

$$ln(\frac{w_{St}}{w_{Ut}}) = \beta_0 + \beta_1 ln(\frac{a_t}{b_t}) + \beta_2 ln(\frac{S_t}{U_t}) + \gamma X \tag{3.4}$$

where $X$ is a vector of control variables and $\gamma$ is a vector of corresponding regression parameters. Alternatively, other researchers treat relative share of skilled workforce as dependent variable and let it depend on wage ratio. This is applied for example by Li (2008) or Li (2010), who estimates the following equation instead of (3.4):

$$ln(\frac{S_t}{U_t}) = \beta_0 + \beta_1 ln(\frac{a_t}{b_t}) + \beta_2 ln(\frac{w_{St}}{w_{Ut}}) + \gamma X \tag{3.5}$$

Interesting twist of the estimation strategy has been proposed by Behar (2010). He suggests to derive elasticity from the estimated parameters as $\sigma = \beta_2 + 2$. This modification prevents from getting extreme values of elasticity if the estimated coefficient is close to 0. Given the fact that no other researcher adopts this approach, we will not present its derivation.

Researchers assuming translog cost function can derive parameters of the function, including the elasticity of substitution between heterogeneous labor from cost share equations. Being linear in parameters, they are convenient to use. In three factor case with skilled labor, unskilled labor and capital, this means to estimate the system of equations (2.16) - (2.18) presented in Chapter 3 of this thesis. This approach is adopted by Bergström & Panas (1992), who simultaneously estimate cost function and cost share equations by iterative Zellner-efficient procedures.

## 3.2 Definition of Skill

Another issue linked with the estimation of the elasticity is how to define and measure skill. In practice, skill level is not observed, therefore appropriate proxies have to be used. There are two streams in the reviewed literature. In the first case, researchers choose to define workers with an education level above certain cut-off value as skilled, other workers are considered unskilled. Different levels of education have been chosen as cut-off points by researchers. For instance, Dupuy & Marey (2007) estimate elasticity between college educated workers and other workers. Ciccone & Peri (2005), on the other hand, consider all workers with at least secondary education as skilled and Mello (2011) uses primary education as a cut-off point.

The other stream identifies skilled and unskilled workers based on type of occupation. Typically, production workers are considered to be unskilled and non-production workers skilled. Skilled and unskilled workers are frequently referred to as "white collars" and "blue collars". An example of the use of this proxy for skill can be found in (Reshef 2007), (Kearney *et al.* 1997) or (Berndt & Christensen 1974).

## 3.3 Long-run and Short-run Elasticity

Most researchers in the reviewed literature do not explicitly declare which type of elasticity they aim to estimate. In some cases, definition of short-run and long-run elasticity used by researchers is rather vague. For instance, Ciccone & Peri (2005) who estimate elasticity using five decennial censuses claim that they "refer to this estimate as the long-run elasticity because estimation relies on 10-year changes in the relative supply of more educated workers and their relative wage." However, to generate their estimates referred to as long-run, they use the same specification as Katz & Murphy (1992), whose estimate based on annual data is in their words "probably best interpreted as short-run substitution elasticity."

For the purpose of our analysis, we only count estimated elasticity as short-run if it is produced using first-difference estimation or error-correction model. Regressions in these forms allow to separate long-term effects and emphasize short-term fluctuations. According to (Behar 2010), first differences is "arguably more appropriate for estimating short-run effects." If we adopt this definition of short-run estimates, only few estimates are assumed to capture short-run substitution effect. For instance, some of the estimates reported by Behar (2010) and by (Mollick 2008) are generated using two-step error-correction model introduced by Engle & Granger (1987). This method consists of including lagged residuals from baseline specification into equation in first-difference form. Alternatively, Dupuy (2007), who borrows models from Katz

& Murphy (1992) and Krusell *et al.* (2000) estimates these models in first-difference form.

## 3.4 Nature of the Data

Reviewed studies also differ in type of data used to produce estimates of the elasticity. One choice researchers have to face is that of data aggregation. Hamermesh (1996) classifies empirical studies of labor demand into three main groups based on level of data aggregation. First, there are studies using aggregated data, where the unit of observation is either an entire economy or large industry. Second, there are studies where small industries or sectors are observed. Third group consists of studies where firms, establishments or individuals are used as units of observation. Ideally, level of aggregation should depend on what precisely is being estimated. In case that researchers are interested in "typical" firm or worker, a natural choice would be to apply micro-level data. If instead, they aim to study labor as aggregate and determine macroeconomic parameters of production function, they need aggregated data. However, there are several potential problems of aggregation. For instance, Hamermesh (1996) mentions problem of linear aggregation of nonlinear relationships. The idea is that even with the assumption of identical technologies in all firms or establishments, one can not expect that parameters in estimated equations are the same for particular firm and for the aggregated case. Moreover, aggregating workers into groups means to implicitly assume that these workers are "very close p-substitutes or q-complements" and that they are equally productive. Another practical issue with aggregated data is that fewer observations used in regressions are usually linked with lower precision. Stanley (2005), for instance, claims that it is possible to use sample size or its square root as a measure of precision[2].

Closely linked to the level of aggregation, another choice for researchers is that of time series data versus cross-sectional data. In some cases, time series are needed. As discussed above, to generate estimates interpretable as short-run, using first-differences and error-correction model requires that time dimension is also reflected in the data. In case that long-run elasticity is estimated, Hamermesh (1996) claims that " (...) there is nothing inherently more attractive in cross-sections or time-series data. Rather, the choice depends

[2]This measure of precision can be used as an alternative to the inverse of standard error when funnel plots are constructed. This is discussed in more detail in Chapter 5.

on the degree of spatial aggregation in each type of available data." In practice, time series on micro-level are quite rare and cross-sectional data generally enable greater disaggregation.

## 3.5 Endogeneity Problem and Instruments

A major issue in empirical literature aiming to estimate substitution elasticity and production function parameters in general is that of potential endogeneity bias. As explained by (Behar 2007), frequently estimated equations treat production inputs as exogenous, producing endogenous output. But it is reasonable to expect that the choice of inputs can be made according to desired output that maximizes profit. In this case, factor inputs are in fact endogenous. This may cause biased econometric estimates of the parameters of production function. Researchers frequently address this problem by instrumenting labor supply and instead of OLS, they use two-stage least squares (2SLS). This approach is restricted by availability of the data: a good instrument has to be uncorrelated with the error term (exogenous) and highly correlated with relative labor supply (relevant). An example can be found in Ciccone & Peri (2005), who use state and year specific compulsory school attendance and child labor laws as instruments for relative labor supply of more educated workers. Silva *et al.* (2007), on the other hand, instruments relative share of educated labor with a variable called "minimum wage intensity" defined as share of workers potentially affected by new minimum wage legislation.

# Chapter 4

# Data Collection and Description

## 4.1 Data Collection

To generate our summary, we collected 684 estimates from 78 studies. These estimates were found using Google Scholar. As a primary searching query, we used:

```
‘‘Elasticity of substitution between skilled and unskilled’’ AND estimate
```

Considering different definitions of skilled labor, we also used queries to search for studies with tertiary education as a cut-off point:

```
‘‘Elasticity of substitution between college and high school’’ AND estimate
```

and for studies using occupation type as skill measure:

```
‘‘Elasticity of substitution between"  and "blue and white collars"
```

For all searching results we went through their abstracts and often through the text to identify empirical studies that generate their own estimates of elasticity rather than assuming value based on theory and previous studies.[1] While collecting existing evidence, we had to be careful and avoid multiple inclusion of studies that are published more than once. Failed identification of duplicate studies may occur, if they present same evidence under differently named paper with different first author. Multiple inclusion of publications in meta-analyses may lead to overweighting duplicate results. Although this might seem trivial,

[1]High number of studies returned by our searching queries were in fact theoretical or estimated other parameters calibrating some value of the elasticity of substitution between skilled and unskilled workers.

examples of meta-analysis with multiple inclusion can be found, as discussed by Thornton & Lee (2000).

Additional studies were found via references in previously found studies. Also, we included most estimates from studies collected and cited in brief reviews of empirical literature by Freeman (1986), Hamermesh (1996) and recently by Behar (2010) that we were able to find and access.[2] We finished our search in September 2017. Complete list of collected studies is provided in Appendix A.

The oldest study we found was published in 1970, four newest studies are from 2017. Almost all collected studies except one are in English, the only exception is study by Jamet (2005) written in French. Studies use data from different countries, most often from United States, Latin American countries or they use large panels of various countries with different characteristics.

Negative estimates of the elasticity of substitution are not excluded, for two reasons. First, although they are not consistent with commonly used definition of the elasticity, they are still interpretable if the estimated effect is Allen-Uzawa elasticity of substitution (AES). Second, even if the estimated elasticity is not of the type allowing for negative values, excluding these estimates could cause additional bias. There is a simple reason for this. As explained by Havranek *et al.* (2017), when there is no upper limit to balance zero threshold, excluding negative estimates and including large positive ones may cause upward bias of the statistics calculated from collected estimates. Although negative estimates are not interpretable *per se*, overall inference may be less precise without them.

Apart from values of elasticity and their standard errors, we collected several other variables useful for the analysis of heterogeneity between existing estimates. These include characteristics of studies themselves, of the data used, estimation methods and strategies, publication characteristics and number of citations. Besides elasticity estimates and standard errors, our dataset contains following variables:

- **year_study**: year of publication, in case of unpublished studies, year when the study first appeared (for instance as a working paper)
- **midyear_data**: middle year of the data used in primary studies in case

[2] There are some studies mentioned in previous literature reviews that we could not find or access. This was mainly caused by the fact that these studies were unpublished dissertations (Especially in case of studies mentioned by Hamermesh (1996)).

researchers used time-series or panel data, year of data collection for cross-sectional data

- **yearcount_data**: number of years captured by the data used in primary studies. In almost all cases, this corresponds to the length of the time-series or panel, sometimes only subset of data corresponding to one year is used to calculate the estimated elasticity. In case of cross-sections coded as 1.

- **n_citat**: number of citations of the primary study on Google Scholar.

- **n_obs**: number of observations used in regression fitted to estimate elasticity.

- **R2**: R-squared or Adjusted R-squared from regression fitted to estimate the elasticity.

- **country**: country or group of countries where the data used in primary study were collected. This categorical variable had too many levels whose occurrence was low. So in final dataset, rare levels are clustered together resulting in three levels: country_US stands for United States, country_dev stands for developing country or group of developing countries, country_other corresponds to all other options including panels of miscellaneous countries.

- **est**: estimation method chosen to estimate regression to obtain elasticity. Again, rare levels were grouped together and resulting categories are: est_2SLS for Two stage least square regression, est_OLS for Ordinary least squares estimation.

- **data**: level of aggregation of the data used in the primary study. As in Hamermesh (1996), there are three levels: data_micro for micro-level data, where units are single workers or firms, data_sector for data on sectoral level and data_aggr for aggregated data, where units are whole economies or large industries.

- **pf**: production function assumed by researchers. Four levels are coded: pf_CES1 as one-level CES function, pf_CESM as multilevel CES function, pf_TL as translog function and pf_XNA for all cases where assumptions about production functions are not disclosed.

- **freq**: frequency of data collection for data in primary study. There are four levels: freq_m_q_s four data with frequency higher than annual which means monthly, quarterly and semi-annually, freq_a for annual collection frequency, freq_more for lower frequency than annual, typically every three, five years or decennial frequency and freq_XNA in case that frequency does not make sense, in case that data is cross-sectional.

- **skill**: definition of skilled and consequently for unskilled workforce used in primary study. If researchers used education as proxy for skill and tertiary education is used as threshold, coded level is skill_col. If researchers used secondary education as threshold corresponding level is skill_high. If researchers used occupation type as proxy for skill, we assign level named skill_occ. In all other cases, skill_other is used.

- **def**: definition of elasticity measure. Three levels are defined - def_HES for Hicks elasticity of substitution, def_AES for Allen-Uzawa elasticity of substitution and def_other for other elasticity measures, such as Shadow elasticity of substitution or Morishima elasticity of substitution.

- **Xsection**: dummy variable coded 1 for cross-sectional data and 0 otherwise.

- **Longrun**: dummy variable that is a proxy of long-run elasticity. We coded 0 in case that researcher estimated error-correction model or used regression in first-difference form, 1 otherwise.

- **Published**: dummy for studies published in scientific journal. As a benchmark, we take group of studies that are in form of working papers, theses etc.

- **impact_factor**: REPEC discounted recursive impact factor - score from ranking of economic scientific journals.[3] For journals not on the list we code 0.

- **Male_only**: dummy for estimated elasticities, for which only male workers were considered.

- **Manufacturing**: dummy for estimates where data reflect labor markets of the manufacturing sector.

[3] Available at: `https://ideas.repec.org/top/top.journals.rdiscount.html`

- **v_time**: dummy for estimates from regression, where time-related variables were included.
- **v_educ**: dummy for estimates from regression, where education-related variables were included as control variables.
- **v_location**: dummy for estimates from regression, where dummies for location were included as regressors.
- **v_macro**: dummy for estimates from regression, where macroeconomic indicators were included as regressors.
- **v_sector**: dummy for sector-related controls.
- **v_age**: dummy for controls related to age of workers aiming to capture additional source of heterogeneity.
- **v_ethn**: dummy for ethnicity, nationality and immigration-related controls.
- **v_capital**: dummy for capital-related controls.

## 4.2 Description of Collected Elasticity Estimates

Table 4.1 presents summary statistics of all collected estimates. We can see that collected estimates are relatively dispersed. They vary from minimum value of -436.85 (0.1) reported by Blankenau & Cassou (2011) to maximum estimate of 1000 (0.111) by Psacharopoulos & Hinchliffe (1972). We can see that median estimated elasticity is equal to 1.4, close to the estimate of 1.41 by Katz & Murphy (1992), the most cited estimate in the literature and often used for calibration. Sample mean is considerably higher, approximately 3.96. Standard deviation is also very high, it is more than ten times higher than the mean. Apparently, large difference between mean and median is driven by few extreme positive estimates that are only partially balanced by extreme negative estimates.[4]

Not all estimates that has been collected may be used for all parts of our analysis. For diagnostics of potential publication selection bias, standard errors

[4] Apart from the estimated elasticity of substitution of 1000 (0.111) by Psacharopoulos & Hinchliffe (1972), there is an estimate equal to 566.08 (0.2) by Blankenau & Cassou (2011), 202 (0.1) by Bowles (1970) and 100 (0.05) by Card (2009). Such extreme negative values are only two - that of Blankenau & Cassou (2011) and -265.39 (0.13) by Blundell *et al.* (2016).

Table 4.1: Elasticity Estimates, Whole Sample

| Statistic | N | Mean | Median | St. Dev. | Min | Max |
|---|---|---|---|---|---|---|
| elasticity | 684 | 3.956 | 1.4 | 49.407 | −436.85 | 1000 |
| elasticity, winsorized | 684 | 1.996 | 1.4 | 2.41 | −0.99 | 11.45 |

are needed for both formal regression-based tests and graphical test. For this reason, in this part of analysis, a simple inclusion criteria is applied: study has to report standard errors of the estimated elasticity of substitution. Searching the existing literature, we found that researchers often fail to report standard errors of their estimates (or t-statistics that can be used to calculate standard errors). Therefore from all collected studies, only a smaller subset is included. Results of the analysis of publication selection only concern 581 out of 684 estimates. Table 4.2 presents summary statistics of the included estimates. As we can see, mean and median value of the elasticity estimates were not largely affected by excluding studies that did not report standard errors. Both minimum and maximum values of collected elasticities remain in the subsample. Consequently, standard deviation of the estimated elasticity in the subsample is even higher than in the whole sample. Table 4.2 also shows summary statistics for standard errors of the elasticity estimates. Ranging from 0.0004 to 534.33, standard errors are also dispersed.

Table 4.2: Elasticity Estimates, Subsample: Non-missing Standard Errors

| Statistic | N | Mean | Median | St. Dev. | Min | Max |
|---|---|---|---|---|---|---|
| elasticity | 581 | 4.177 | 1.42 | 53.541 | −436.85 | 1000 |
| elasticity, winsorized | 581 | 1.962 | 1.42 | 2.43 | −0.99 | 11.45 |
| std.error | 581 | 1.568 | 0.183 | 23.372 | 0.0004 | 534.33 |
| std.error,winsorized | 581 | 0.293 | 0.183 | 0.336 | 0.006 | 1.673 |

To better illustrate distribution of the existing estimates, Figure 4.1 shows histogram of the empirical evidence of labor substitutability. For the sake of readability, histogram is constructed omitting 15 most extreme values of the elasticity. Histogram illustrates how the distribution of the estimates is positively skewed. Relatively high number of estimates falls between zero and two. There are numerous estimates with a value lower than 1, providing some evidence for complementarity of the two types of labor. There are even more estimates above 1, implying that skilled and unskilled workers are to some extent substitutable, but far from being perfect substitutes. Histogram omits 15 most extreme estimates including extremes from both tails of the distribution.

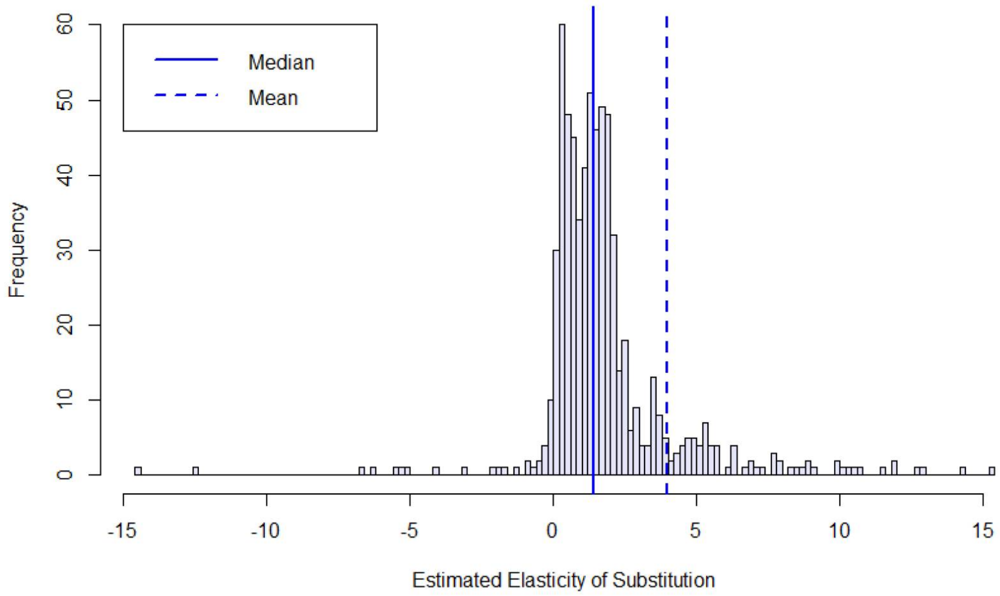



Figure 4.1: Histogram of Estimated Elasticities

*Figure shows the histogram of collected estimates of the elasticity of substitution between skilled and unskilled labor. For better readability, 15 most extreme estimates were excluded resulting in N=669. Mean and median values are calculated from all estimates including extremes.*

Before we proceed to diagnostics of publication selection bias, we briefly discuss values of other collected variables. Table 4.3 shows descriptive statistics for numeric variables. Table 4.4 further explores categorical variables, shows number of estimates corresponding to each category and mean and median of elasticity in subsample corresponding to each category. Looking at these summary statistics, we can get a better idea about how the estimates are distributed based on different attributes of the collected studies. However, any impetuous conclusion based on simple statistics might be misleading. To correctly interpret existing empirical evidence of labor substitutability, we have to address outliers, unbalanced number of estimates per study, potential publication bias and separate the effects of study and estimate characteristics using combination of regression coefficients.

Table 4.3: Descriptive Statistics for Numeric Variables

| Statistic | N | Mean | St. Dev. | Min | Max |
|---|---|---|---|---|---|
| year_study | 684 | 2002 | 11.58 | 1970 | 2017 |
| midyear_data | 684 | 1982 | 12.15 | 1929 | 2010 |
| yearcount_data | 684 | 20 | 15.63 | 1 | 91 |
| n_citat | 684 | 90.63 | 288.79 | 0 | 4941 |
| n_citat_py | 684 | 5.03 | 14.97 | 0 | 190.04 |
| n_obs | 466 | 781.95 | 5060.01 | 9 | 72321 |
| impact_factor | 344 | 0.921 | 0.138 | 0.00 | 5.04 |
| R2 | 428 | 0.57 | 0.3 | 0.000 | 0.99 |

## 4.3 Adjustments to Collected Data

Information provided by Table 4.1 and Table 4.2 shows that values of the estimated statistics may be dominated by small number of outliers. Some of the primary studies reported extreme values of elasticity or standard errors. These outliers do not have to be omitted, as this would lead to complete loss of information provided by these observations. Instead, the effect of potentially spurious outliers is addressed using winsorization. With this technique, extreme values are not simply omitted, but they are replaced based on value of chosen percentile leading to adjusted, "winsorized" values:

$$\sigma_w = min(max(\sigma, \sigma_p), \sigma_{(1-p)}) \tag{4.1}$$

$$se(\sigma)_w = min(max(se(\sigma), se_p), se_{(1-p)}) \tag{4.2}$$

where $p$ is percentage of data we want to replace from each tail, $\sigma_p$ and $se_p$ are computed percentiles representing lower thresholds for elasticity estimates and their standard errors and $\sigma_{(1-p)}$ and $se_{(1-p)}$ stand for corresponding upper thresholds (Miller 1993). In our case, we calculated $2.5^{th}$ and $97.5^{th}$ percentiles for elasticity (These are equal to -0.99 and 11.45, respectively) and standard error (0.006 and 1.673, respectively). Table 4.1 and 4.2 show summary statistics for winsorized elasticity for the whole sample and subsample with non-missing standard errors. Moreover, Table 4.2 also reports these standard errors after winsorization. It can be seen that winsorization has substantial effect on sample mean and standard deviation. Mean winsorized elasticity is approximately 1.996 for the whole sample and 1.962 for subsample with non-missing standard errors as opposed to original values of 3.956 and 4.177. Standard deviation

Table 4.4: Categorical Variables, Elasticity Statistics by Category

| *Variable* | $N_{total}$ | $N_1$ | $N_0$ | *Mean elasticity* | *Median elasticity* |
|---|---|---|---|---|---|
| country_US | 684 | 244 | 440 | 2.20 | 1.55 |
| country_dev | 684 | 214 | 470 | 1.16 | 0.98 |
| country_other | 684 | 226 | 458 | 8.5 | 1.38 |
| est_2SLS | 684 | 154 | 579 | 3.02 | 1.69 |
| est_OLS | 684 | 105 | 530 | 12.85 | 1.84 |
| est_other | 684 | 425 | 259 | 0.97 | 1.12 |
| data_micro | 684 | 81 | 604 | 3.19 | 1.9 |
| data_sect | 684 | 300 | 384 | 2.35 | 0.69 |
| data_aggr | 684 | 304 | 380 | 5.74 | 1.85 |
| pf_CES1 | 684 | 399 | 258 | 0.99 | 1.25 |
| pf_CESM | 684 | 156 | 519 | 4.31 | 1.64 |
| pf_TL | 684 | 89 | 595 | 1.96 | 1.28 |
| pf_xna | 684 | 31 | 653 | 45.92 | 2.23 |
| freq_m_q_s | 684 | 51 | 633 | -1.44 | 2.99 |
| freq_a | 684 | 397 | 287 | 2.65 | 1.16 |
| freq_more | 684 | 67 | 617 | 1.14 | 1.54 |
| freq_xna | 684 | 169 | 515 | 9.77 | 1.45 |
| skill_col | 684 | 197 | 487 | 4.78 | 1.92 |
| skill_high | 684 | 123 | 561 | 10.8 | 1.85 |
| skill_occ | 684 | 347 | 337 | 1.20 | 0.73 |
| skill_other | 684 | 17 | 667 | 0.94 | 0.83 |
| def_HES | 684 | 590 | 94 | 4.30 | 1.42 |
| def_AES | 684 | 72 | 612 | 2.03 | 1.31 |
| def_other | 684 | 22 | 662 | 1.01 | 0.8 |
| Xsection | 684 | 169 | 515 | 9.77 | 1.45 |
| Longrun | 684 | 633 | 51 | 4.02 | 1.38 |
| Published | 684 | 344 | 340 | 7.06 | 1.74 |
| Male_only | 684 | 111 | 573 | 4.68 | 1.92 |
| Manufacturing | 684 | 277 | 407 | 1.02 | 0.64 |
| v_time | 684 | 415 | 269 | 2.10 | 1.24 |
| v_educ | 684 | 16 | 668 | 2.04 | 1.42 |
| v_location | 684 | 80 | 604 | 1.32 | 1.61 |
| v_macro | 684 | 34 | 650 | 1.38 | 1.98 |
| v_sector | 684 | 21 | 663 | 0.47 | 0.37 |
| v_age | 684 | 40 | 644 | 2.10 | 1.24 |
| v_ethn | 684 | 11 | 673 | 1.91 | 1.43 |
| v_capital | 684 | 57 | 627 | 2.51 | 1.61 |

shrank from approximately 50 to near 2.4 in both whole sample and subsample of estimates. Winsorized standard errors have mean value of 0.293 with standard deviation 0.336.

Typically, several estimates are taken from a single study. The number of estimates taken from each study varies greatly. From just one estimate up to as much as 185 different estimates estimated by Reshef (2007)[5]. This would give much more weight to this study compared to studies with lower number of reported estimates. This might be problematic especially if there is unobserved heterogeneity between studies. To address unbalanced number of estimates in study-level clusters, regressions can be weighted using inverse of the number of estimates reported by studies. For example, in case of $j^{th}$ estimate from study $i$ that reports $n_i$ estimates, weight is defined as $w_{ij} = \frac{1}{n_i}$. This weighting scheme is used in our analysis of heterogeneity between studies. As a robustness check, it is also used in tests for publication bias. But in regressions for testing publication bias, our preferred weighting scheme is to use precision as weights, for two reasons. First, it has straightforward interpretation as funnel asymmetry test. Second, it is assumed to be a better corrective measure for heteroskedasticity. More detailed explanation is provided in the next chapter.

[5]Large number of estimates is caused by the fact that Reshef (2007) estimates elasticities across different industries for three countries using three different estimation techniques.

# Chapter 5

# Addressing Publication Bias

## 5.1 Definition of Publication Selection Bias

In previous chapter, we described substantial evidence about the elasticity of substitution between the two skill groups that we have collected. However, it is practically impossible to gather all existing scientific knowledge about any phenomenon, for one simple reason: not all findings are reported. As Dickersin (1990) pointed out, although science depends on reporting, there are no commonly accepted standard for decisions about reporting. This might be problematic in case that results with certain characteristics have systematically lower probability of being reported. Then the evidence that gets reported (and published) may be a biased representation of existing knowledge. This phenomenon is called publication selection bias or publication selectivity problem. In a broader sense, it does not only concern published studies. Bias may be present also in unpublished evidence in form of working papers of unpublished theses. But there is a difference in its potential source: unpublished estimates are not subject to judgement of publishers, selectivity in case of unpublished studies reflects reporting decision of researchers. As a part of evidence is "hidden in the file drawers", selection bias is sometimes referred to as file-drawer problem (Rosenthal 1979).

What are the characteristics of studies and estimates that determine the probability of publication? Ideally, studies should be selected for submission and publication based on their quality. This type of selectivity is desirable and it is not harmful by itself. On the other hand, if studies are selected based on their results, bias may occur. Typically, researchers and publishers have tendency not to submit and publish studies that generate values that are not consistent with theory and previous empirical literature (Thornton & Lee 2000). This type of preferential selectivity based on magnitude or direction of the estimated effect is called type I publication selection bias. Moreover, studies are less likely to be published if they yield so-called "negative results", in other words, if they fail to reject null hypothesis of nonexistence of the studied effect. An evidence of such bias was presented by an early study by Smart (1964), who reviewed psychological research. Regardless of their direction, studies whose results are not statistically significant are also less likely to be selected. This is called type II bias (Stanley 2005). These are most common sources of publication selectivity, but not the only sources that has been discussed.

Various other causes of publication bias are discussed by Thornton & Lee (2000) including: (*i*) faulty design and confusing reporting method of single studies, (*ii*) selective publication based on sponsorship, (*iii*) faulty design and execution of reviews and meta-analyses. As for (*iii*), Thornton & Lee (2000) mention three potential sources of additional bias created in reviews: selectivity based on language of the study, multiple inclusion of results and wrong choice of regression models used in meta-regression analysis. As reviewers, we have to be careful not to create additional bias. Inevitably, our collection of estimates is to some extent non-random given the fact that we search for and review articles written in English [1]. Even if we included all studied written in languages we understand, sample would be incomplete. Moreover, this selective pattern is only harmful if English studies systematically differ in magnitude, sign and significance from studies written in other languages.

There are multiple meta-analyses that previously identified publication bias in existing economic literature. For instance, type I bias is detected by Havranek & Irsova (2011) in empirical evidence of vertical spillovers from FDI or by Doucouliagos *et al.* (2014) in evidence concerning income elasticity of a statistical life. Stanley (2005), tested for bias of type II in 73 published studies estimating productivity effects of unionization and found and detected signs of selectivity based on statistical significance.

Selective publication may lead to wrong conclusions about the true values of estimated parameters in the population. For instance, sample median and sample mean of published estimates may be biased estimates of the true effect. This occurs if published studies systematically differ from unpublished studies. One straightforward implication of the existence of publication bias is the dan-

[1] With an exception of (Jamet 2005) written in French and found through a reference in another reviewed study.

ger that results of meta-analyses may be invalidated in case that publication selection bias exists and is not treated properly. For an empirical assessment of the effect of publication bias on meta-analyses, see Sutton *et al.* (2000). Therefore in attempt to summarize existing empirical evidence, we have take into account a possibility that some part of knowledge about studied phenomenon may be underreported.

## 5.2 Testing for Publication Selectivity in Reviewed Literature

### 5.2.1 Funnel Plot

There is a simple and commonly used graphical method to identify potential publication selection bias in reviewed literature. Funnel plot is a scatterplot of estimated values on horizontal axis plotted against their estimated precision on vertical axis. If there is no publication selection bias in the existing literature, funnel plot should be symmetrical around the "true" value (Duval & Tweedie 2000). In practice, it is assumed that most precise estimates are close to the true value, thus we will be interested in symmetry around the most precise estimates. Moreover, as the most precise estimates are expected to be close to the true value, plot should have a shape of an inverted funnel.

There are at least two proxies for precision used to construct funnel plots in meta-analytic studies: the inverse of standard errors of the collected estimates and number of observations used in primary studies to generate these estimates (Stanley 2005). We use inverse of standard errors as proxy for precision. To use alternative measure of precision - number of observations - is in this case suboptimal, for two reasons. Many primary studies do not report number of observations used in their regressions. Moreover, data used in primary studies differ in their level of aggregation. This complicates interpretation of differences between the number of observations used in regressions by reviewed studies.

To our best knowledge, there is not yet any study that contains funnel plot of existing estimates of the elasticity of substitution between skilled and unskilled workers. We thus plot collected estimates against their precision and resulting funnel plot is reported in Figure 5.1. It has to be said that these are not only studies published in academic journals or books, but also working papers and dissertations. Nevertheless, the interpretation is similar. Only the estimates with reported standard errors or t-values can be plotted. For better readability, observations with the most extreme precision and estimated elasticity are not on the plot. Vertical lines represent median and mean value of elasticities including extreme estimates that are not on the plot. As we can see in Table 4.2, median and mean are approximately 1.42 and 4.18, respectively. Median value being close to the most precise estimates indicate that estimates are relatively evenly divided into two groups - the group of the estimates that are on the left from the most precise estimates and those that on the right. Mean is shifted to the right from the most precise estimates. But as we discussed in the previous chapter, it is largely driven by few extreme outliers that are not plotted.

Constructed funnel plot is not exactly shaped as inverted funnel. There are more values with relatively high precision on the right side of the funnel plot. On the left side, estimates are mostly close to the most precise estimates. Fewer estimates are actually negative and they are rather imprecise. This can point out to the fact that negative estimates can be systematically underreported.

However, if there is heterogeneity between existing estimates, interpretation of the funnel plot becomes less straightforward. In presence of heterogeneity, funnel plot can be asymmetric even when there is no bias. Typically, such funnel plots have many relatively precise values far from the assumed true value which creates multiple peaks. In Figure 5.1, we can see that there are some more precise estimates with value of elasticity between 7.5 and 10. Though less elevated, there appear to be second peak. Although it may be that precision of these estimates is overestimated, this can point out to the fact that there are genuinely more true effects depending on what exactly is measured.

Figure 5.2 shows funnel plots of the two subsets of collected estimates - those that were published in academic journals and those that were not. Funnel plot of published estimates is wider, contains higher number of negative values. Fewer negative estimates in the subset of unpublished studies might be interpreted as a slight positive bias in unpublished studies. Interestingly, the most precise estimates are in the subset of unpublished estimates. Also, most unpublished estimates are close to the assumed true value. As for possible evidence of heterogeneity, second peak visible in Figure 5.1 is mostly driven by published estimates, though unpublished estimates close to 10 are also to some extent more precise. To conclude, based on graphical analysis of funnel plots, there seem to be some evidence of publication bias. It is mostly due to collected estimates from unpublished working papers and dissertations. The

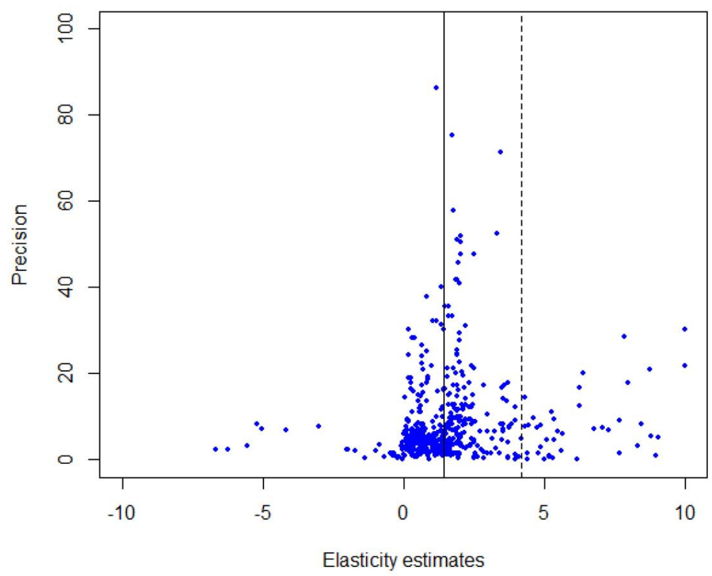



Figure 5.1: Funnel Plot of Estimated Elasticities

*Figure shows the funnel plot of collected estimates of the elasticity of substitution between skilled and unskilled labor. Solid vertical line represents sample mean and dashed vertical line represents median of the elasticity estimates.16 most extreme observations are not plotted resulting in N=534. But all estimates are included in regressions.*

asymmetry of funnel plot for unpublished estimates seems to be more severe than it is for estimates published in academic journals.

Nevertheless, graphical methods that aim to detect publication selection bias are still subject to the interpretation of the reviewer. As suggested by Stanley (2005), they are not sufficient by themselves as the "symmetry may be in the eye of the beholder". Also, as explained above, potential heterogeneity may complicate interpretation of funnel's skewness.

### 5.2.2 Regression-based Tests for Publication Bias

There exist some more rigorous alternatives to graphical methods of detection of publication selection bias. These techniques are based on regression analysis. First specification used for this purpose has been presented by Card & Krueger (1995). According to Stanley (2005), it is based on assumption that in the absence of publication bias, estimates should be randomly distributed around the true effect and independent of their standard error. The estimated equation can be written as:

$$\sigma_{ij} = \beta_0 + \beta_1 \cdot se(\sigma_{ij}) + \epsilon_{ij} \tag{5.1}$$

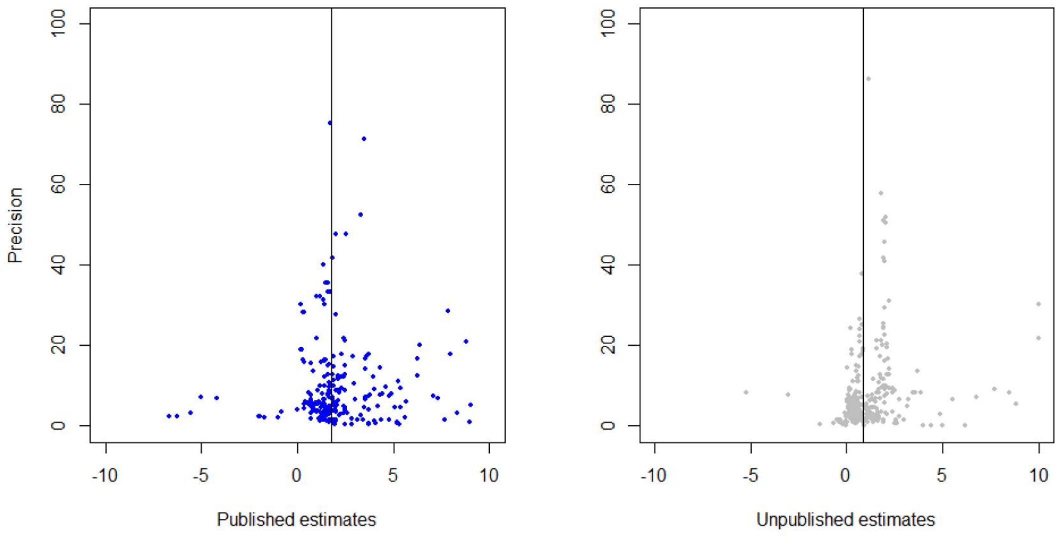



Figure 5.2: Funnel Plot of Estimated Elasticities: Published and Unpublished

*Figure shows the funnel plot of collected estimates of the elasticity of substitution between skilled and unskilled labor. Solid vertical line represents median of the elasticity estimates. Most extreme observations are not plotted resulting in N=236 and N=294 for published and unpublished estimates, respectively. But all estimates are included in regression.*

where, in our case, $\sigma_{ij}$ stands for $i^{th}$ estimate of the elasticity of substitution between skilled and unskilled labor from $j^{th}$ empirical study, $\sigma_{ij}$ stands for its reported standard error and $\epsilon_{ij}$ for the error term. If the estimated coefficient $\beta_1$ is significantly different from zero, there is an indication that publication selection bias is present in the literature. Sign and magnitude of the estimated coefficient is interpretable as direction and extent of this bias. Testing for statistical significance of $\beta_0$, on the other hand, is in general interpretable as the test of the existence of true effect different from zero. In our case, if we fail to reject null hypothesis of $\beta_1$ being equal to zero, this would imply zero elasticity between the two groups of workers.

Usually, it is assumed that $\epsilon_{ij}$ is independently and identically distributed: $\epsilon_i | se(\hat{\sigma}_{ij}) \sim N(0, \sigma^2)$. In our case, this is probably violated. There is no reason expect that residuals for observations linked to the same study will be independent, nor can we expect that they are identically distributed if they are reported by different studies. In the presence of such unknown correlation between errors, the estimates may still be unbiased. The problem is that default standard errors are likely to overstate precision which may lead to wrong inference based on statistical tests. One option available with sufficiently large number of clusters[2] is to fit regressions assuming clustered standard errors. This way, we explicitly allow for correlation between errors within the same cluster (i.e. linked with the estimates reported by the same study) and assume independence across clusters (studies). Mathematically written, condition $E(\epsilon_{ij}|se(\sigma_{ij})) = 0$ becomes $E(\epsilon_{ij}, \epsilon_{i'j'}|se(\sigma_{ij}), se(\sigma_{i'j'})) = 0$ unless $j = j'$. Therefore our first method used to estimate equation (5.2) is ordinary least squares (OLS) with standard errors clustered at the study level.

Secondly, we test for publication selection bias by estimating model with study-level fixed effects (FE), also with clustered standard errors. Using fixed effects on the study level, we control for unobserved heterogeneity between the studies. In fact, the estimated equation becomes:

$$\sigma_{ij} = \beta_0 + \beta_1 \cdot se(\sigma_{ij}) + \alpha_j + \eta_{ij}, \eta_{ij} \sim N(0, \theta) \tag{5.2}$$

where $\epsilon_{ij}$ gets decomposed into two components: $\alpha_j$ captures study-level fixed effects and $\eta_{ij}$ stands for estimate-level disturbances. Fixed-effects model is allowing for correlation between $\alpha_j$ and standard error of the $i^{th}$ estimate from the $j^{th}$ study, written $E(\alpha_j|se(\hat{\sigma}_{ij})) \neq 0$.

As an alternative to fixed-effects model, we also estimate multilevel mixed-effects model (ME). Within this framework, error component $\alpha_j$ is now representing study-level random effects. It is treated as uncorrelated with our regressor - standard error of the estimated elasticity, written $E(\alpha_j|se(\hat{\sigma}_{ij})) = 0$. Study-level random effects are assumed to follow normal distribution with mean 0 and variance $\zeta$ conditional on $se(\hat{\sigma}_{ij})$: $\alpha_j \sim N(0, \zeta)$.

First, we run these regressions for all collected estimates that have reported standard errors. Then we inspect potential bias in subsamples of published and unpublished studies. Unlike funnel plots, regressions also take into account observations with extreme precision and extreme values of estimated elasticity, these outliers are addressed using winsorization. Results from unweighted regressions are reported in Table B.1 in Appendix B. The results are consistent with our interpretations of funnel plots. Based on these unweighted regressions, there seem to be some evidence of publication selectivity. For the pool of all estimates, only Mixed effects estimation yields statistically significant estimate of publication bias. Moreover, it seems to be driven by unpublished studies.

[2]According to Cameron & Miller (2015), consensus number of clusters believed to be "sufficiently close to infinity" is 50 for balanced panels, probably even higher for unbalanced panels. We have 62 clusters with unbalanced number of observations within clusters.

Unweighted versions of Fixed effects and Mixed effects estimations point out to positive bias in these studies.

However, unweighted regressions are most likely suboptimal. Equation (5.1) is very likely to suffer from heteroskedasticity. It is expected that more extreme estimates are linked with higher standard error. This expectation is supported by graphical evidence from funnel plot. Consequently, estimated standard errors can be wrong which leads to incorrect conclusions about statistical significance. A common way to correct for heteroskedasticity is to use weighted least squares. As weights, meta-analysts (see, for instance Egger *et al.* (1997), Stanley (2005) or Havranek *et al.* (2017)) often weight specification (5.1) by the inverse of standard error of the estimate:

$$\frac{\hat{\sigma}_{ij}}{se(\hat{\sigma}_{ij})} = \beta_1 + \beta_0 \cdot \frac{1}{se(\hat{\sigma}_{ij})} + \epsilon_i \tag{5.3}$$

When used as weight, inverse of the standard error is assumed to be a measure of the heterosedasticity of residuals in (5.1). Also, with this weighting scheme, precise estimates are given more weight than less precise ones. In (5.3), the dependent variable is in fact t-statistic of the elasticity estimate. $\beta_1$ becomes intercept and $\beta_0$ represents slope. However, the interpretation of these coefficients is still similar to the previous case: testing for the statistical significance of $\beta_1$ can be interpreted as test for publication bias. With this specification assumed, the test is called funnel asymmetry test (FAT). Statistical significance of $\beta_0$ points out to the fact that the "true effect" is present, which in our case means that the elasticity is not zero. This test is called precision effect test (PET) (Stanley 2005). Results from regressions estimating (5.3) are presented on the left side of Table 5.1. Results from regressions estimating this equation on subsamples of published and unpublished estimates are presented on the left side of Table 5.2.

As a robustness check, we use alternative weighting scheme. Instead of using precision as weight and fitting (5.3), we use specification (5.1) weighting regressions by the inverse of the number of estimates per study. This way, every study is getting the same weight, studies with large number of estimates are penalized. Results from these alternative regressions are presented on the right side of Table 5.1 and Table 5.2.

For all regressions, estimates of the constant are statistically insignificant on all conventional levels. Interestingly, although the magnitude of estimates is not negligible, its sign is not consistent for all the estimations. Coefficients

Table 5.1: Regression-based Tests for Publication Selectivity Bias, Weighted

| | *Dependent variable: estimated elasticity* | | | | | |
|---|---|---|---|---|---|---|
| | *All estimates* | | | | | |
| | PRECISION | | | STUDY | | |
| | OLS Clustered | FE Clustered | ME | OLS Clustered | FE Clustered | ME |
| Std.Error | 3.641 | -6.814 | 4.839 | 0.070 | -0.803 | 0.070 |
| (Bias) | (9.404) | (5.479) | (10.200) | (0.527) | (1.200) | (0.532) |
| Constant | 2.662*** | 3.358*** | 3.177*** | 2.542*** | 2.798*** | 2.542*** |
| (Effect) | (0.484) | (0.365) | (0.147) | (0.364) | (0.352) | (0.367) |
| Observations | 581 | 581 | 581 | 581 | 581 | 581 |
| Studies | 62 | 62 | 62 | 62 | 62 | 62 |
| F-statistic | 29.16 | 84.7 | | 0.02 | 0.45 | |
| Wald Chi-sq | | | 469.19 | | | 0.02 |

*Note: Estimated elasticities and their standard errors are winsorized.*

Table 5.2: Continued: Regression-based Tests for Publication Selectivity Bias, Weighted

| | *Dependent variable: estimated elasticity* | | | | | |
|---|---|---|---|---|---|---|
| | *Unpublished estimates* | | | | | |
| | PRECISION | | | STUDY | | |
| | OLS Clustered | FE Clustered | ME | OLS Clustered | FE Clustered | ME |
| Std.Error | 0.754 | -6.683 | 5.901 | 0.127 | 1.589 | 0.127 |
| (Bias) | (5.607) | (7.863) | (10.849) | (0.998) | (0.936) | (1.018) |
| Constant | 1.922*** | 2.465*** | 2.179*** | 2.162*** | 1.772*** | 2.159*** |
| (Effect) | (0.051) | (0.575) | (0.213) | (0.579) | (0.249) | (0.591) |
| Observations | 318 | 318 | 318 | 263 | 263 | 263 |
| Studies | 26 | 26 | 26 | 26 | 26 | 26 |
| F-statistic | 1342.77 | 18.41 | | 0.03 | 2.88 | |
| Wald Chi-sq | | | 104.50 | | | 0.02 |
| | *Published estimates* | | | | | |
| | PRECISION | | | STUDY | | |
| | OLS Clustered | FE Clustered | ME | OLS Clustered | FE Clustered | ME |
| Std.Error | 7.536 | 5.025 | 11.706 | -0.034 | -2.362 | -0.034 |
| (Bias) | (8.999) | (0.420) | (13.951) | (0.585) | (1.506) | (0.619) |
| Constant | 3.373*** | 3.524*** | 3.429*** | 2.843*** | 3.573*** | 2.843*** |
| (True Effect) | (0.325) | (0.420) | (0.205) | (0.455) | (0.472) | (0.474) |
| Observations | 263 | 263 | 263 | 263 | 263 | 263 |
| Studies | 36 | 36 | 36 | 36 | 36 | 36 |
| F-statistic | 102.42 | 70.27 | | 0.00 | 2.46 | |
| Wald Chi-sq | | | 279.04 | | | 0.00 |

estimated by fixed effects estimation are negative, unlike coefficients from clustered OLS and mixed effects estimation. The size and sign of this coefficient can be interpreted as the estimate of size and direction of publication selection bias, respectively. Therefore, using regression-based tests, we did not find significant and consistent evidence of publication bias in literature as a whole. On the other hand, regression coefficients of the inverse of the standard error are statistically significant for all the regressions. As we mentioned above, testing for the significance of this coefficient can be interpreted as testing for nonzero true effect. Being statistically significant and ranging from approximately 2.54 to 3.36, estimate of the true value of elasticity is significantly different from zero and above 2.

Table 5.2 shows the estimates for unpublished and published studies separately. In case of unpublished studies, results are similar to results of regressions for all studies. There are no significant evidence for publication bias. Estimated bias from OLS and mixed effects estimation is positive, while the estimate from fixed effects estimation is negative. However, robust standard errors are too large to confirm presence of bias in any direction. Estimated true effect is between 1.77 and 2.47 and significantly different from zero. For published studies, estimated true effect is higher than for unpublished studies - from 2.84 in case that OLS is used to as much as 3.52 for fixed effects estimation. True effect is significant on 1 % level of significance in all cases. Again, the hypothesis of the absence of publication bias cannot be rejected on any conventional level, despite the fact that estimated coefficients are relatively large in magnitude and consistent in their sign.

Results from regressions using inverse of number of estimates as weights are consistent with our findings following from use of specification (5.3). All regressions yield statistically insignificant estimates of publication selection bias. Based on weighted regressions that we believe to be superior to unweighted regressions, we do not reject null hypothesis of no publication bias in the existing empirical literature. This is in contrast with our interpretation of funnel plot, which lead us to expect slight positive bias. Apart from possible heterogeneity, funnel asymmetry might be driven by few studies that report higher number of estimates with unconventional combination of elasticity value and precision. For instance, out of 10 estimates close to 10 that report relatively high precision, 5 are from the same study by Verdugo (2014). In Appendix B, we show funnel plot of mean elasticities and mean precisions reported by each study. Despite low number of points, resulting funnel plot is relatively symmetrical around the most precise values.

# Chapter 6

# Drivers of the Heterogeneity

## 6.1 Methodology

To determine what drives differences between the existing estimates of the elasticity of substitution between skilled and unskilled labor, we apply method called Bayesian model averaging (BMA). What is our motivation to choose this approach instead of commonly used frequentist meta-regression analysis?

First, we have many regressors. To be specific, we wish to inspect 32 explanatory variables. Every variable is either included in the chosen specification, or not. As a result, number of possible specifications reaches $2^{32}$. With so many regressors, we face substantial uncertainty about which model is optimal to explore heterogeneity of the estimates. Chosen model could be easily underfitted or overfitted. Obviously, to manually explore all possibilities by estimating each model is practically impossible. Even if we used automatic model selection procedure, such as stepwise regression, we would still need to properly address model uncertainty, as suggested by Chatfied (1995). Also Madigan & Raftery (1994) claim that to ignore model uncertainty might be suboptimal and suggest model averaging, as it usually allows better predictions than choosing a single model.

Second, Bayesian framework allows a more natural interpretation provided by the data than Frequentism, especially in case of small datasets and in case that it is not possible to repeat sampling many times. VanderPlas (2014) explains:

> "Bayesian approaches (...) are often conceptually more straightforward, and pose results in a way that is much closer to the questions a scien-

tist wishes to answer: i.e. how do these particular data constrain the unknowns in a certain model?"

In our case, question is: Which characteristics of studies and particular estimates have impact on the value of elasticity? Frequentist approach would be to test for significance of the explanatory variable which would yield p-values. This is interpretable as the probability of getting "such extreme" data given the null is true. Bayesian approach, on the other hand, allows us to estimate probability that this variable explains part of the variation of the response variable given the collected data. Moreover, as suggested by Kass & Raftery (1995), non-bayesian significance testing was developed mainly to be used for comparison of two models (typically nested), but in our case, comparison of more than two models is needed.

Bayesian model averaging consists of fitting large number of models determined by available set of explanatory variables and calculating weighted average of these regressions (Zeugner 2011). Weights are constructed based on application of Bayes theorem and they are called Posterior Model Probabilities (PMP). Let $D$ be sampled data, $N$ the number of possible explanatory variables, $K$ be number of possible models[1] such as set of possible models is defined as $M = \{M_1, M_2, ..., M_K\}$. Then posterior model probability of model $M_k$, where $k = 1, 2, ..., K$ is defined as:

$$pr(M_k|D) = \frac{pr(D|M_k)pr(M)}{pr(D)} = \frac{pr(D|M_k)pr(M)}{\sum_{k=1}^{K} pr(D|M_k)} \tag{6.1}$$

Here, numerator is the product of marginal likelihood of model $M_k$ and prior model probability. Denominator sums up all marginal likelihoods for models from set $M$. Marginal likelihood of the model $M_k$is defined as:

$$pr(D|M_k) = \int pr(D|\beta_k, M_k)pr(\beta_k|M_k)d\beta_k \tag{6.2}$$

where $\beta_k$ stands for a vector of regression parameters associated with model $M_k$, $pr(\beta_k|M_k)$ is prior density of $\beta_k$ and $pr(D|\beta_k, M_k)$ is likelihood in conventional form. Prior model probability is a formulation of prior beliefs about probability distribution of possible models (Hoeting *et al.* 1999).

The goal is to find the best possible approximation of the distribution of regression parameters conditioned upon sampled data and set of models. To get

[1]Note that in general, $K = 2^N$, which in this case results in $N = 32$ and $K = 2^{32}$

an idea of how this distribution looks like, reported output from BMA contains the set of three statistics - weighted posterior mean, weighted posterior variance and posterior inclusion probability - for every considered independent variable. Weighted posterior mean estimates the effect of this variable on the target, its interpretation is thus comparable to the interpretation of coefficient in linear regression. For $i^{th}$ variable, it is defined as :

$$E(\beta_i|D) = \sum_{k=1}^{K} \hat{\beta}_{ik} pr(M_k|D) \tag{6.3}$$

Here, $\hat{\beta}_{ik}$ is the estimated regression coefficient for the $i^{th}$ variable estimated by model $M_k$. Posterior variance then can be written as:

$$Var(\beta_i|D) = \sum_{k=1}^{K} (Var(\beta_i|D, M_k) + \hat{\beta}_{ik}) pr(M_k|D) - E(\beta_i|D)^2 \tag{6.4}$$

Posterior inclusion probability (PIP) is defined as the sum of posterior model probabilities for all models where the $i^{th}$ candidate regressor is included (Zeugner 2011). This statistic is reported as an alternative to frequentist statistical significance. Perhaps the most controversial point of the implementation of Bayesian framework is that assumptions about prior probabilities have to be made. Implementation of BMA is not an exception. Equation (6.1) requires to formulate beliefs about prior model probability $pr(M)$. Moreover, as we can see from equation (6.2), assumptions about prior density of parameters are also needed. In the next section, we discuss our choice of priors and reasoning behind this choice.

In practice, implementation of BMA is computationally demanding, for two reasons. Hoeting *et al.* (1999) explained that the number of terms in the above equation is often very high which implies that summation needed to calculate posterior probability distributions might be problematic. Also, integrals needed to calculate (6.2) are often difficult to compute. There are two ways to manage summation. First approach is called Occam's window and it was introduced by Madigan & Raftery (1994). It consists of defining subset of models using two criteria. It first discredits all models that have far worse predictive power that the best model with the threshold set by the researcher. Then it applies Occam's razor principle to exclude models that are complex and have less support from sampled data that simpler alternatives. The second approach is to use Markov Chain Monte Carlo (MCMC) sampling method. Markov chain is a stochastic process with memoryless property (Hermanns 2002). MCMC uses Markov Chain to repeatedly sample from probability distribution without having much prior information about this distribution. In spite of its stochastic nature, it is a powerful tool allowing to solve deterministic problems. When applied to our model averaging problem, using Metropolis-Hastings algorithm, one initial model from the pool of possible models is chosen randomly. Then at each iteration, a new candidate model (say, $M_b$) is proposed against the previously sampled model ($M_a$). The algorithm follows by calculating the so-called acceptance ratio, defined as the ratio of probabilities of the two models at a given iteration. Then a new model is chosen against previous one with the following probability:

$$p_{a,b} = min(1, p(M_b|D)/p(M_a|D)). \tag{6.5}$$

Information about rejection and acceptance is stored. Number of times each model is accepted over another model is used to construct posterior model probability (Hoeting *et al.* 1999) (Zeugner 2011). For a more detailed theoretical discussion of computational algorithms used to implement BMA, see (Raftery *et al.* 1997) or (Draper 1995).

## 6.2 Implementation

We wish to estimate:

$$\hat{\sigma}_{ij} = \beta X_{ij} + \epsilon_{ij}, \quad \epsilon_{ij} \sim N(0, \theta) \tag{6.6}$$

where $\hat{\sigma}_{ij}$ is the $i^{th}$ estimate from the $j^{th}$ study, $X_{ij}$ is a vector of explanatory variables concerning estimate itself and study where the estimate has been reported. $\beta$ is corresponding vector of regression coefficients, and $\epsilon$ is a normally distributed error term. We first run BMA procedure. Results from BMA are confronted with regression coefficients and their standard errors from OLS with clustered standard errors. As a further robustness check, we run BMA assuming different priors. Vector of estimated coefficients $\hat{\beta}$ from BMA is not from a single model, but is equal to posterior mean determined by combination of large number of models.

Set of possible collected variables is presented in Chapter 4. Because of too many missing values for variables *n_obs* (number of observation used in estimation regression or regressions) and *R2* (R-squared from estimation regression) we exclude these variables from our analysis because too many rows would be automatically dropped in BMA. Moreover, some researchers report adjusted R-squared and others R-squared. One of these measures penalizes overfitting, the other does not. Similarly, the fact that level of data aggregation differs between studies greatly reduces comparability of the reported number of observations. There is another variable that has many missing values - *impact_factor*. This is expected, because of inclusion of unpublished studies. We decide to code 0 for these observations instead of dropping the variable. Each categorical variable has been coded into set of dummies, now we have to drop one dummy for each of these variables (to avoid dummy variable trap).

To prevent multicollinearity, before running BMA procedure we construct correlation matrix for all variables that we intend to use for heterogeneity analysis. Correlation matrix is reported in Appendix C. Figure C.1 visualizes all correlations between the variables. Red color stands for positive correlation, blue points to negatively correlated variables. Table C.1 shows all pairs of variables whose correlation is higher than 0.6 or lower than -0.6. The most correlated variables are *def_AES* and *pf_TL* with correlation of almost 0.9. It is due to the fact that most researchers who assume translog production function also estimate Allen-Uzawa elasticity of substitution. Since these two variables carry similar information about dependent variable and potential multicollinearity could cause imprecise estimates that are very sensitive to minor changes in model, we decide to drop variable *pf_TL*. Final number of explanatory variables used for heterogeneity analysis is thus 32.

Variables can be divided into following subcategories: *Real Factors* are variables that can possibly influence true value of elasticity, not just the estimates. Group *Elasticity Type* contains variables that show which definition of the elasticity is used - HES, AES or other (typically Morishima or Shadow) and if the relationship is long-run or short-run. Definition of the production or cost function is closely linked to the definition of elasticity, thus variables concerning choice of production function are also in this group. *Data* specify what type of dataset is used in primary study, if it is cross-sectional or not, its level of aggregation, frequency of data collection, proxy for skill. Category *Estimation* captures researcher's choices of estimation method and precise specification including choice of controls. Last group of variables, *Publication*, contains publication characteristics and number of citations. We choose to use number of citations per year instead of total number of citations because we do not want to penalize recent studies.

We run BMA using R package BMS by Zeugner (2011). Our baseline calibration of priors and sampling properties is as follows:

*(i) mprior='random'*

This option allows to incorporate our expectations about model prior probability. Setting model prior to *random* means that we assume distribution of the model size to be beta-binomial. This is a conservative assumption that reflects no prior knowledge about model distribution and that no model size is preferred. We use this prior following the example of Ley & Steel (2009). Alternatively, we set model prior to *uniform* to see if the results are sensitive to a change of model prior.

*(ii) g='BRIC'*

Here, $g$ stands for constant $g$ in Zellner's $g$ prior. For multivariate normal linear regression model, Zellner's $g$ prior can be used as a reference prior distribution for regression coefficients. It is a multivariate normal distribution that has covariance matrix specified depending on the data (Wang 2015). In our case, Zellner's $g$ prior can be written as: $\beta|\theta \sim N(\beta_0, g\theta(X'X)^{-1})$. $\beta_0$ is usually assumed to be zero as size and magnitude of the coefficients is not *a priori* known. Choice of constant $g$ can impact significantly the resulting inference. *BRIC* stands for prior suggested by Fernandez *et al.* (2001) who argued that higher g minimizes prior impact on the results. It is defined as $max(N, K^2)$, where $N$ is the number of observations and $K$ stands for number of candidate regressors Zeugner (2011). In our case, it is equal to 1024. Our alternative choice, *UIP* sets g to number of observations (N) in our case equal to 684.

*(iii) mcmc='bd'*

Here, *mcmc* stands for Markov Chain Monte Carlo sampling which is used to construct posterior distributions. Within this framework, there exist multiple samplers. That is, there are more possibilities of how to choose candidate model in each iteration. *bd* stands for birth-death sampler. According to Zeugner (2011), this is the standard choice for BMA procedure. This sampler proposes new candidate model the following way: out of all possible variables, sampler draws one variable randomly. If this variable is in previous model, it will be dropped, if not, it will be added.

*(iv) iter=2 000 000*

We choose 2 million iterations to ensure convergence.

*(v) burn=1 000 000*

The first model that is randomly chosen by the sampler may not be an ideal starting point for good approximation of the posterior distribution. However, after some while sampler tends to converge to better models. That is why not all the results are stored to compute posterior probabilities. *Burn* stands for number of iterations that are not stored, also called *burn-ins*. Number of burn-ins is set to 1 million.

*(vi) nmodel=5 000*

*Nmodel* stands for the number of best models from which information will be stored and reported. Only small share of models is highly informative [2]. Given the fact that choosing too many models significantly slows down sampler, we choose only 5000 best models.

## 6.3 Results from BMA

Results from our baseline BMA procedure are reported in the left panel of Table 6.1 and in Figure 6.1. Further details including posterior densities for important variables are provided in Appendix D.

The left panel of Table 6.1 shows estimated posterior means (Post Mean), posterior standard deviation (Post StD.) and posterior inclusion probabilities (PIP) for all variables. Economic interpretation of posterior mean and posterior standard deviation is similar to the interpretation of regression coefficient and its estimated standard error in simple regression. But what about the interpretation of PIP? What probability is considered to be sufficiently large to infer significant relationship between the regressor and dependent variable? We follow Kass & Raftery (1995) and Havranek *et al.* (2017) and interpret PIP according to following heuristic. Variables with posterior probabilities higher than 0.5 carry significant information about variation of the dependent variable. Moreover, PIP between 0.5 and 0.75 implies "weak significance", PIP between 0.75 and 0.95 points out to "positive significance ", values between 0.95 and 0.99 and higher than 0.99 imply "strong" and "decisive" significance, respectively.

[2]Typically, most models have very small posterior probabilities. This can be seen on Figure 6.2.

Table 6.1: Heterogeneity of Elasticity estimates

| | *Results from BMA* | | | *Frequentist Check: OLS* | | |
|---|---|---|---|---|---|---|
| | Post Mean | Post StD. | PIP | Coeff. | Std. Error | P-value |
| *Real Factors* | | | | | | |
| country_dev | -1.371 | 0.237 | 1.000 | -1.393*** | 0.408 | 0.001 |
| country_US | -0.927 | 0.260 | 0.980 | -0.918** | 0.393 | 0.019 |
| midyear_data | -0.075 | 0.018 | 0.982 | -0.082*** | 0.026 | 0.002 |
| manufacturing | 1.552 | 0.436 | 0.985 | 1.475*** | 0.554 | 0.008 |
| male_only | 0.001 | 0.036 | 0.025 | | | |
| *Elasticity type* | | | | | | |
| def_AES | -0.049 | 0.214 | 0.074 | | | |
| def_other | -1.391 | 0.787 | 0.836 | -1.539* | 0.906 | 0.089 |
| longrun | 2.291 | 0.442 | 1.000 | 2.231** | 0.922 | 0.015 |
| pf_xna | -1.256 | 0.279 | 0.999 | -1.154 | 1.045 | 0.270 |
| pf_CESM | 0.015 | 0.080 | 0.059 | | | |
| *Data* | | | | | | |
| data_micro | -0.073 | 0.203 | 0.154 | | | |
| data_sect | -0.021 | 0.112 | 0.055 | | | |
| Xsection | -0.761 | 0.460 | 0.801 | -0.946 | 0.597 | 0.113 |
| yearcount_data | 0.007 | 0.013 | 0.311 | | | |
| freq_m_q_s | 0.211 | 0.467 | 0.214 | | | |
| freq_more | 0.032 | 0.146 | 0.069 | | | |
| skill_high | 3.353 | 0.312 | 1.000 | 3.237*** | 1.001 | 0.001 |
| skill_col | 0.176 | 0.275 | 0.344 | | | |
| *Estimation* | | | | | | |
| est_2SLS | 0.117 | 0.295 | 0.174 | | | |
| est_OLS | 0.137 | 0.056 | 0.165 | | | |
| v_time | -2.186 | 0.237 | 1.000 | -2.120*** | 0.711 | 0.003 |
| v_capital | -1.870 | 0.297 | 1.000 | -1.942*** | 0.549 | 0.000 |
| v_age | 1.649 | 0.375 | 0.999 | 1.757** | 0.292 | 0.000 |
| v_location | -1.126 | 0.763 | 0.799 | -1.279** | 0.579 | 0.027 |
| v_educ | -0.130 | 0.284 | 0.217 | | | |
| v_sector | 0.163 | 0.576 | 0.122 | | | |
| v_ethn | -0.035 | 0.207 | 0.050 | | | |
| v_macro | 0.012 | 0.103 | 0.040 | | | |
| *Publication* | | | | | | |
| published | 1.170 | 0.178 | 1.000 | 1.158** | 0.503 | 0.021 |
| year_study | 0.075 | 0.018 | 0.982 | 0.082*** | 0.026 | 0.001 |
| impact_factor | -0.017 | 0.057 | 0.118 | | | |
| n_citat_py | 0.000 | 0.001 | 0.058 | | | |
| (Constant) | -0.017 | | 1.000 | -0.013 | 0.029 | 0.644 |

*Notes: N=684. Weighted by inverse number of estimates per study. Elasticity is winsorized. OLS std.errors are clustered on study level. Mprior = random, g= BRIC.*

According to the estimates, there is some evidence that characteristics of the estimates and studies are related to the estimated elasticity of substitution between skilled and unskilled labor. In each group of our candidate regressors, there are several variables that significantly explain variation in elasticity estimates.

*Real Factors*
From all collected variables, the most important factor having influence on the value of elasticity is the country or countries whose labor markets are reflected in the data. Our evidence suggests lower elasticity of substitution for developing countries and for United States. Baseline category contains all other countries including European countries, Canada, Australia, New Zealand and estimates from large panels of various countries with different characteristics. Both categories - *country_US* and *country_dev* - have PIPs that imply significance and posterior means that are relatively large in magnitude (-0.927 and -1.371, respectively). Time appears to be another important " real factor". With PIP reaching 0.982 and negative posterior mean, middle year of the data is also strongly significant and suggests that the elasticity of substitution may slightly diminish in time.[3] On the contrary, if elasticity is measured for manufacturing sector only, the elasticity estimates tend to be significantly higher.

*Elasticity type*
Definition of the elasticity measure also appears to affect resulting estimate. According to BMA results, holding other things constant, long-run elasticity estimates are higher. This is consistent with the theory that usually assumes long-run elasticities to be higher. Interestingly, although Allen-Uzawa elasticity can have negative values (as opposed to Hicks elasticity), Allen-Uzawa elasticities are not linked to significantly lower elasticity according to BMA. Posterior mean os *def_AES* is negative, but PIP is relatively low. Instead, compared to Hicks elasticity, other definitions, such as Morishima or Shadow elasticity are related to significantly lower elasticity. Regarding choice of production function, we found no link between the value of elasticity and use of multilevel CES form instead of one-level CES. However, researchers that do not explicitly report their choice of production function tend to report lower probabilities, as the estimate of dummy *pf_xna* is strongly significant and negative.

[3]However, we are aware of the fact that midyear of data is not an ideal proxy to capture development of elasticity in time. This variable captures only middle year, not the starting and ending points in time. It is thus only approximate measure of time. It is used because of heterogeneous nature of datasets in primary studies.

*Data*
Elasticity estimates also seem to be sensitive on certain characteristics of the data. Regarding data characteristics, choice of proxy for skill is the most important factor. Compared to the baseline choice - occupation type - researchers choosing education with high school as threshold seem to get significantly higher estimates. This variable has very high PIP, 1.000 which can be interpreted as "decisive" significance. Another factor that seems to explain variation in estimated elasticity is the choice of cross-sectional data. According to negative BMA posterior mean and high PIP, using Cross-sections is linked with lower estimated elasticity. Frequency of data collection does not seem to play an important role when it comes to elasticity values. We found no evidence for sensitivity of elasticity estimate to the level data aggregation.

*Estimation*
Differences in estimation method do not seem to systematically affect elasticity estimates in one direction. This does not mean that choice of method is not important, researchers themselves warn that in some cases, using OLS may lead to biased and inconsistent estimates [4]. What seems to have an effect on resulting elasticity estimate is the choice of control variables by researchers in primary studies. Controls related to time, capital, age and location have all high PIP, ranging from 0.799 to 1.000. All except control for age have negative posterior means.

*Publication*
Last group of variables is related to publication and citation of the primary study. According results from BMA, published studies report higher estimates of the elasticity. Variable *published* has PIP of 1.000 which corresponds to decisive significance and positive posterior mean. Quite unexpectedly, year of publication has positive posterior mean and high PIP. It seems to be in contrast with negative posterior mean of the variable *midyear_data*. The interpretation of negative posterior mean is that newer studies are linked with higher estimates of the elasticity. But this does not mean that the elasticity is growing in time, as the year of publication is not decisive.

The right panel of Table 6.1 is showing results from OLS regression[5] that includes variables with posterior inclusion probability higher than 0.5. We call this regression "Frequentist check". The threshold of 0.5 is not arbitrary,

[4] See, for instance Blankenau & Cassou (2011)
[5] Again, standard errors are clustered at the study level.

but corresponds to lower threshold for significance according to the above-mentioned rule of thumb.

Frequentist check mostly confirms our results from BMA. Most variables that have high posterior probabilities are statistically significant, even with robust standard errors. There are only two exceptions: variables *pf_xna* and *Xsection* are not significant on any conventional level. Signs of coefficients from OLS are consistent with signs of posterior means. Magnitude of the estimates is also similar.

Figure 6.1 shows inclusion of variables in 5000 models with highest posterior model probabilities. This follows from our baseline calibration of BMA procedure. Each column represents one model and models are sorted according to these probabilities. Variables are sorted based on their posterior inclusion probabilities. Blank spaces illustrate the fact that a given variable is not included in corresponding model. Red color means negative sign of regression coefficient of the variable in the model, blue color represents positive sign. Regressions are weighted by inverse of the number of estimates per study. We can see that signs of regression coefficients are consistent across different models which means that our straightforward interpretation of signs of posterior means is not wrong - best models consistently assume same direction for the estimated effects.

As we discussed earlier in this chapter, choice of priors is perhaps the weakest point of Bayesian Model Averaging and can have impact on the results. For this reason, we run BMA with alternative model prior and alternative Zellner's g. Results from alternative BMAs with different choices of model and coefficient priors are reported in Appendix E. We can conclude that our results are fairly robust to change in priors. Same set of variables is assumed to be important for the value of estimated elasticity. There are just small changes in magnitude of the coefficients and values of posterior inclusion probabilities.

## 6.4 Synthetic Estimates: Best Practice Approach

As our final step in quantitative analysis of empirical literature, we use collected estimated elasticities and coefficients from BMA procedure to generate synthetic estimates of the elasticity. This section is inevitably to some extent subjective, as different researchers may have different opinion about how to define best practice.

What is thus our definition of suggested approach? It is preferred to use dis-

aggregated data (or so-called microdata). The reason for this choice is that incorrect aggregation may lead to wrong conclusions about elasticity value. This is discussed in detail in Chapter 3 of this thesis. We would recommend to choose time series over cross-sectional data.[6] Preferred frequency is annual. Higher frequency of data collection creates additional complications for the estimation - researchers might need to address seasonality which is not always the case. As proxy for skill, we choose education level with secondary education as cut-off value. According to Hamermesh (1996), using occupation as proxy for skill and dividing labor to production and nonproduction workers is "a somewhat negative example" of defining groups of labor, because there is a large overlap in earnings between these two groups. Moreover, Hamermesh claims that the estimation of elasticity between these two groups is "of little inherent policy interest".
We are interested in up-to-date estimate, we thus plug in sample maximum for the variable *midyear_data*. As for estimation strategy, given the fact that researchers probably face endogeneity and few studies actually estimate structural equations, 2SLS should be chosen over OLS. We plug in 1 for all dummy variables indicating use of controls in regressions used to estimate elasticity. Also, we plug in sample maxima for impact factor and number of citations per year and 1 for dummy variable indicating that study in which the estimate is reported is published in academic journal.

Based on above defined best practice, we generated multiple synthetic estimates: for US, Developing countries and other countries. In each case we generate both long-run and short-run estimates of both Hicks elasticity of substitution and Allen-Uzawa elasticity of substitution. Results are presented in Table 6.2. Table shows estimates and corresponding 95% confidence intervals. Confidence intervals are calculated based on OLS regression including all variables used in BMA. Our best-practice estimates of long-run elasticity range from 2.44 to 3.87 depending on precise definition of the elasticity and studied countries. These estimates are considerably higher than the estimate 1.41 by Katz & Murphy (1992) - the most cited long-run estimate that is frequently used by researchers calibrating substitution elasticity. Allen-Uzawa elasticity of substitution is consistently slightly lower than Hicks elasticity. Estimated long-run elasticity is lower for United States and for the group of developing countries compared to other countries. Our short-run estimates implied by defined best-practice range from 0.15 for Allen elasticity of substitution in

[6] Assuming that in both cases, microdata are available.

Table 6.2: Estimated elasticity based on Our Definition of Best Practice

| | | | *Estimate* | *95% CI* | |
|---|---|---|---|---|---|
| Long-run | US | AES | 2.88 | 0.56 | 5.20 |
| | | HES | 2.94 | 0.82 | 5.07 |
| | Developing | AES | 2.44 | -0.14 | 5.01 |
| | | HES | 2.50 | 0.15 | 4.85 |
| | Other | AES | 3.81 | 1.37 | 6.24 |
| | | HES | 3.87 | 1.65 | 6.09 |
| Short-run | US | AES | 0.59 | -1.97 | 3.15 |
| | | HES | 0.65 | -1.73 | 3.03 |
| | Developing | AES | 0.15 | -2.66 | 2.95 |
| | | HES | 0.21 | -2.39 | 2.80 |
| | Other | AES | 1.52 | -1.18 | 4.21 |
| | | HES | 1.58 | -0.92 | 4.08 |

developing countries to 1.58 for Hicks elasticity of substitution in benchmark group of countries.

What is the interpretation of these synthetic estimates? Based on estimated coefficients from BMA and our definition of best-practice estimation, skilled and unskilled workers are substitutes in the long-run. In the sort-run, their substitution is more limited. For United States and developing countries, our estimates imply that skilled and unskilled workers are complements in the short-run. If our definition of best-practice is well-defined, these synthetic estimates should be closer to true values than simple averages, as they aim to filter out possible effects of misspecifications.

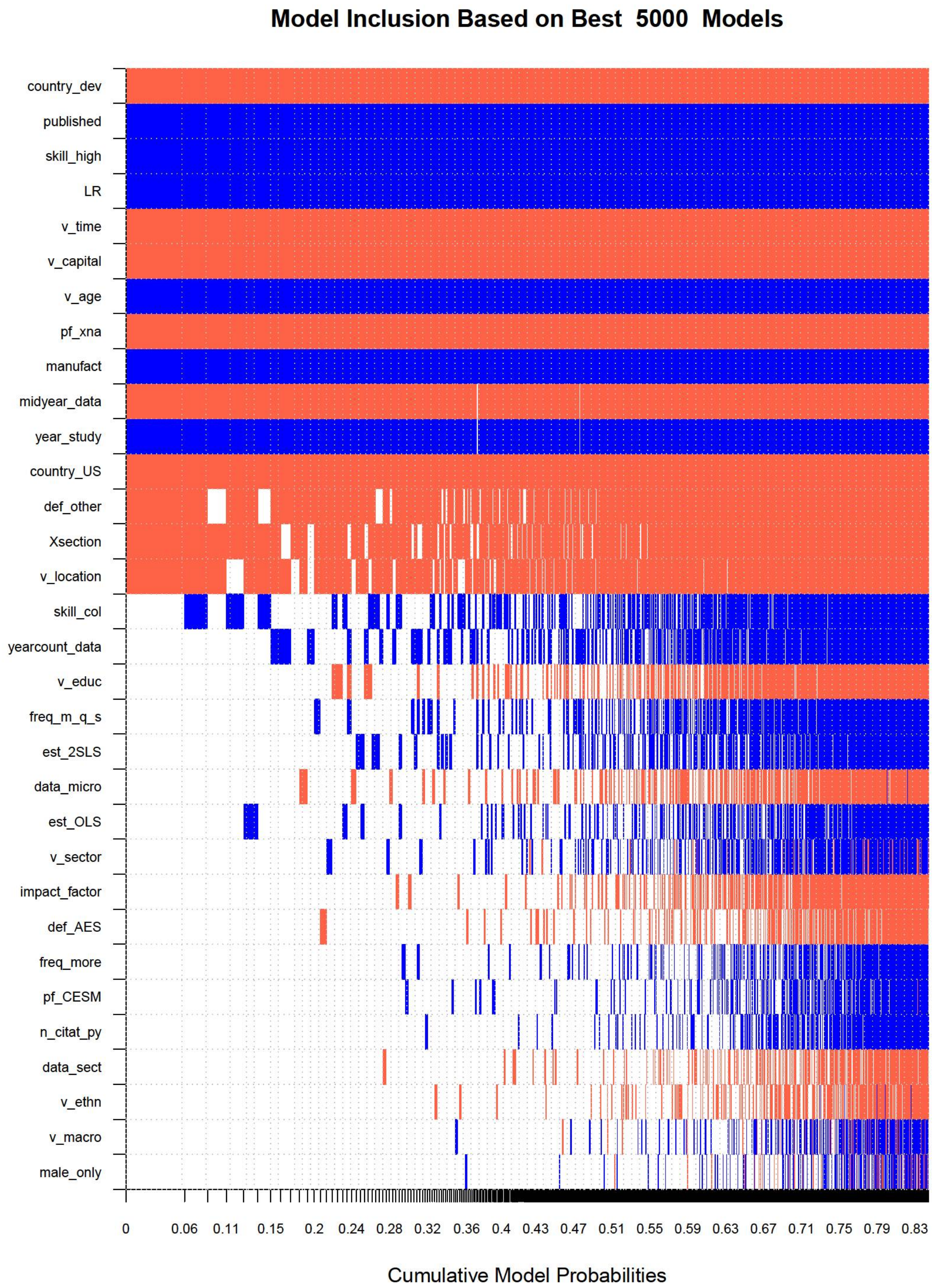



Figure 6.1: Inclusion of explanatory variables in BMA, signs of coefficients

# Chapter 7

# Conclusion

In this thesis, we provide quantitative overview of empirical literature concerning substitution elasticity between skilled and unskilled labor. To summarize existing knowledge about this parameter of the production function, we apply methods of modern meta-analysis. Our overview is based on 684 estimates from 78 studies. These estimates reflect labor substitution in different countries, most frequently United States or group of developing countries. Most estimates show long-run elasticity, but some estimates that were produced using first-difference regressions or error correction model reflect short-run relationship. Existing studies report wide range of estimates including extreme values and negative values not consistent with the theory. These are also included in our analysis to avoid additional bias.

Differences in precise definition of the elasticity used by researchers complicate their comparison. For this reason, we first introduce different estimation strategies and theoretical frameworks behind elasticity estimation. Researchers have multiple options for assumed production or cost function. Some use one-level CES production function, others assume multilevel CES and there are studies using translog framework. There are also differences in measures of substitutability. Most frequently, researches estimate Hicks elasticity of substitution or Allen-Uzawa elasticity of substitution. But there are studies estimating Morishima elasticity of substitution or so-called Shadow elasticity.

Before drawing any conclusions based on collected estimates, we have to test for existence of bias created by publication selectivity. First we construct funnel plots - scatter plots of the estimated elasticities against the inverse of their standard error. Then we provide regression-based tests for publication bias using OLS with study-level clustering of error-term, Fixed effects and

Multilevel mixed effects. Different weighting schemes are applied to correct for heteroskedasticity and unbalanced number of estimates per study. Based on multiple weighted regressions (which we believe to be superior to unweighted regressions), we fail to reject null hypothesis of no publication bias. Despite the fact that negative estimates of Hicks elasticity of substitution are not consistent with theory, they are probably not systematically underreported.

Heterogeneity of reported estimates is examined using Bayesian Model Averaging (BMA). This procedure allows to account for uncertainty in model selection. Large number of candidate regressors - variables that capture characteristics of studies and estimates - imply large number of possible specifications of meta-regression equation. Using BMA, we estimate the effect of these factors on the value of estimated elasticity. Reported parameters are not from a single model, but they are weighted averages of large number of estimated coefficients, weighted by their posterior probabilities.

Based on results from BMA, there are multiple factors that significantly influence value of the estimated elasticity. Some of these factors may reflect real differences in the true value of the elasticity - we call them "real factors". Others reflect research design and estimation strategy or publication characteristics. We find that both real factors and research design influence value of the elasticity. This means that differences between the estimates are partly driven by choices of researchers. However, there are genuinely more true values of the elasticity regardless of the study design. The most important real factor is the country studied. Estimates that reflect US labor market, as well as estimates from developing countries are lower than the estimates from other countries. Time may also play a role - studies with later midyear of the data tend to report lower estimates. In line with the theory, estimates of the long-run elasticity are systematically higher than the estimates of the short-run elasticity.

As for research design and estimation strategy, we find multiple factors that influence resulting value of the elasticity. An important choice of researchers is that of proxy for skill. Choosing education level with secondary education as cut-off point leads to significantly higher estimates of the elasticity. Choosing cross-sectional data over time-series may, on the other hand, contribute to lower estimates. What is also relevant for estimated value is the use of controls in regression equation used to estimate substitution elasticity. Interestingly, although regression-based tests for publication bias do not show the existence of publication bias, BMA implies higher estimates for published studies. To check for robustness of our results we use alternative priors and OLS regression.

These mostly confirm our results - same factors are found to be important for elasticity value.

Estimated regression coefficients from BMA are used to generate new, synthetic estimates. To eliminate the effect of suboptimal choices of researchers on estimated elasticity, we define suggested research design that we call best-practice. It consists of using disaggregated data and time-series with annual frequency of data collection, secondary and higher education as proxy for skill, two-stage least squares estimation with time, capital, sector, location, age, education, ethnicity and economy-related control variables. This definition of best-practice yields long-run elasticity estimates ranging from 2.44 to 3.87 and short-run estimates with values between 0.15 and 1.58. Therefore BMA estimates and best-practice approach imply that skilled and unskilled workers are substitutes in the long-run. In the short-run, their substitutability is more limited, our estimates for United States and for developing countries imply that these two groups of workers are complements in the short-run.

Practitioners who need to calibrate the elasticity face a challenging task to draw conclusions from the existing estimates. The empirical studies yield numerous estimates that differ in sign and magnitude. To collect representative sample of the population of estimates takes considerable amount of time and even if substantial evidence is collected, it might not be sufficient to correctly asses size of the effect. Our contribution is that we provide guidelines for any reader interested in the interpretation of the existing evidence. We hope that the thesis may help to (i) understand different definitions of the elasticity and corresponding estimation strategies, (ii) get an idea about distribution of the existing estimates (iii) understand what drives differences between the estimates, (iv) raise awareness about danger of creating additional bias due to misspecification and inappropriate publication selectivity based on results instead of quality, (v) get additional synthetic estimates of the elasticity.

# Bibliography


Acemoglu, D. (2002): "Technical change, inequality, and the labor market." *Journal of economic literature* **40(1)**: pp. 7–72.

Angrist, J. D. (1995): "The economic returns to schooling in the West Bank and Gaza Strip." *The American Economic Review* pp. 1065–1087.

Askilden, J. E. & Ø. A. Nilsen (2005): "Apprentices and young workers: A study of the Norwegian youth labour market." *Scottish Journal of Political Economy* **52(1)**: pp. 1–17.

Autor, D. H., L. F. Katz, & M. S. Kearney (2008): "Trends in US wage inequality: Revising the revisionists." *The Review of economics and statistics* **90(2)**: pp. 300–323.

Avalos, A. & A. Savvides (2006): "The manufacturing wage inequality in Latin America and East Asia: openness, technology transfer, and labor supply." *Review of Development Economics* **10(4)**: pp. 553–576.

Behar, A. (2007): *Are skilled and unskilled labour complements or substitutes?* Ph.D. thesis, University of Oxford.

Behar, A. (2009): "Directed technical change, the elasticity of substitution and wage inequality in developing countries." .

Behar, A. (2010): "The elasticity of substitution between skilled and unskilled labor in developing countries is about 2." .

Bergström, V. & E. E. Panas (1992): "How robust is the capital-skill complementarity hypothesis?" *The Review of Economics and Statistics* pp. 540–546.

Berndt, E. R. & L. R. Christensen (1973): "The translog function and the substitution of equipment, structures, and labor in US manufacturing 1929-68." *Journal of econometrics* **1(1)**: pp. 81–113.

Berndt, E. R. & L. R. Christensen (1974): "Testing for the existence of a consistent aggregate index of labor inputs." *The American economic review* **64(3)**: pp. 391–404.

Berndt, E. R. & C. J. Morrisson (1979): "Income redistribution and employment effects of rising energy prices." *Resources and Energy* **2(2-3)**: pp. 131–150.

Binelli, C. (2015): "How the wage-education profile got more convex: evidence from Mexico." *The BE Journal of Macroeconomics* **15(2)**: pp. 509–560.

Blankenau, W. F. & S. P. Cassou (2011): "Industry estimates of the elasticity of substitution and the rate of biased technological change between skilled and unskilled labour." *Applied Economics* **43(23)**: pp. 3129–3142.

Blundell, R., D. Green, W. Jin *et al.* (2016): "The UK wage premium puzzle: How did a large increase in university graduates leave the education premium unchanged?" *Technical report*, Institute for Fiscal Studies.

Borghans, L. & B. Ter Weel (2008): "Understanding the technology of computer technology diffusion: Explaining computer adoption patterns and implications for the wage structure." .

Borjas, G. J. (2003): "The labor demand curve is downward sloping: Reexamining the impact of immigration on the labor market." *Technical report*, National Bureau of Economic Research.

Borjas, G. J. & L. F. Katz (2007): "The evolution of the Mexican-born workforce in the United States." In "Mexican immigration to the United States," pp. 13–56. University of Chicago Press.

Bound, J., J. Groen, G. Kezdi, & S. Turner (2004): "Trade in university training: Cross-state variation in the production and stock of college-educated labor." *Journal of Econometrics* **121(1)**: pp. 143–173.

Bowles, S. (1970): "Aggregation of labor inputs in the economics of growth and planning: experiments with a two-level CES function." *Journal of Political Economy* **78(1)**: pp. 68–81.

Broadstock, D., L. Hunt, & S. Sorrell (2007): "UKERC Review of evidence for the rebound effect: Technical report 3: Elasticity of substitution studies." .

Cameron, A. C. & D. L. Miller (2015): "A practitioner's guide to cluster-robust inference." *Journal of Human Resources* **50(2)**: pp. 317–372.

Card, D. (2009): "Immigration and inequality." *Technical report*, National Bureau of Economic Research.

Card, D. & A. B. Krueger (1995): "Time-series minimum-wage studies: A meta-analysis." *The American Economic Review* **85(2)**: pp. 238–243.

Card, D. & T. Lemieux (2001): "Can falling supply explain the rising return to college for younger men? A cohort-based analysis." *The Quarterly Journal of Economics* **116(2)**: pp. 705–746.

Chatfied, C. (1995): "Model uncertainty, data mining and statistical inference (with discussion)." *Journal of the Royal Statistical Society, series B* **158**: pp. 419–466.

Choi, K.-S., J.-H. Jeong, & J. Hwa Jung (2005): "The rising supply of college graduates and declining returns for young cohort: The case of Korea." *Global Economic Review* **34(2)**: pp. 167–180.

Ciccone, A. & G. Peri (2005): "Long-run substitutability between more and less educated workers: evidence from US states, 1950–1990." *Review of Economics and statistics* **87(4)**: pp. 652–663.

Corker, R. & T. Bayoumi (1991): "Apartheid, growth and income distribution in South Africa: Past history and future prospects." .

Cruz, M., E. Milet, & M. Olarreaga (2017): "Online exports and the wage gap." .

Das, G. G. (1999): "What is assumed in the GTAP database's disaggregation of labor by skill level?" .

David, H. *et al.* (2014): "Skills, education, and the rise of earnings inequality among the "other 99 percent"." .

Denny, M. & M. Fuss (1977): "The use of approximation analysis to test for separability and the existence of consistent aggregates." *The American Economic Review* pp. 404–418.

Dickersin, K. (1990): "The existence of publication bias and risk factors for its occurrence." *Jama* **263(10)**: pp. 1385–1389.

Doucouliagos, H., T. Stanley, & W. K. Viscusi (2014): "Publication selection and the income elasticity of the value of a statistical life." *Journal of Health Economics* **33**: pp. 67–75.

Dougherty, C. R. (1972): "Estimates of labor aggregation functions." *Journal of Political Economy* **80(6)**: pp. 1101–1119.

Draper, D. (1995): "Assessment and propagation of model uncertainty." *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 45–97.

Dupuy, A. (2007): "Will the skill-premium in the Netherlands rise in the next decades?" *Applied Economics* **39(21)**: pp. 2723–2731.

Dupuy, A. & P. Marey (2007): "Shifts and twists in the relative productivity of skilled labor." .

Duval, S. & R. Tweedie (2000): "Trim and fill: A simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis." *Biometrics* **56(2)**: pp. 455–463.

Egger, M., G. D. Smith, M. Schneider, & C. Minder (1997): "Bias in meta-analysis detected by a simple, graphical test." *Bmj* **315(7109)**: pp. 629–634.

Engle, R. F. & C. W. Granger (1987): "Co-integration and Error correction: Representation, estimation, and testing." *Econometrica: journal of the Econometric Society* pp. 251–276.

Fallon, P. R. & P. R. G. Layard (1975): "Capital-skill complementarity, income distribution, and output accounting." *Journal of political economy* **83(2)**: pp. 279–302.

Fernandez, C., E. Ley, & M. F. Steel (2001): "Benchmark priors for Bayesian model averaging." *Journal of Econometrics* **100(2)**: pp. 381–427.

FERNANDEZ SIERRA, M. & J. MESSINA (2017): "Skill premium, labor supply and changes in the structure of wages in Latin America." .

FITZGERALD, J., I. KEARNEY *et al.* (2000): "Convergence in living standards in Ireland: the role of the new economy." *Technical report*.

FÖLDVÁRI, P. & B. VAN LEEUWEN (2006): "Is there an optimal allocation of skills? A two-level CES approach." .

FREEMAN, R. B. (1975): "Overinvestment in college training?" *Journal of human resources* pp. 287–311.

FREEMAN, R. B. (1986): "Demand for Education." *Handbook of labor economics* **1**: pp. 357–386.

FREEMAN, R. B. & J. L. MEDOFF (1982): "Substitution between production labor and other inputs in unionzed and nonunionized manufacturing." *The Review of Economics and Statistics* pp. 220–233.

GALLEGO, F. A. (2012): "Skill premium in Chile: studying skill upgrading in the South." *World Development* **40(3)**: pp. 594–609.

GANCIA, G. A., A. MÜLLER, & F. ZILIBOTTI (2011): "Structural development accounting." .

GIANNARAKIS, P. (2015): "Can job displacements explain the UK's productivity puzzle?" .

GLITZ, A. & D. WISSMANN (2017): "Skill premiums and the supply of young workers in Germany." .

GOLDIN, C. D. & L. F. KATZ (2009): *The race between education and technology*. Harvard University Press.

GYIMAH-BREMPONG, K. & A. O. GYAPONG (1992): "Elasticities of factor substitution in the production of education." *Economics of Education Review* **11(3)**: pp. 205–217.

HAMERMESH, D. S. (1996): *Labor demand*. Princeton University Press.

HAVRANEK, T. & Z. IRSOVA (2011): "Estimating vertical spillovers from FDI: Why results vary and what the true effect is." *Journal of International Economics* **85(2)**: pp. 234–244.

HAVRANEK, T., M. RUSNAK, & A. SOKOLOVA (2017): "Habit formation in consumption: A meta-analysis." *European Economic Review* **95**: pp. 142–167.

HECKMAN, J. J., L. LOCHNER, & C. TABER (1998): "Explaining rising wage inequality: Explorations with a dynamic general equilibrium model of labor earnings with heterogeneous agents." *Review of economic dynamics* **1(1)**: pp. 1–58.

HERMANNS, H. (2002): "Markov Chains." In "Interactive Markov Chains," pp. 35–55. Springer.

HICKS, J. (1932): *The theory of wages.* London, Macmillan.

HOETING, J. A., D. MADIGAN, A. E. RAFTERY, & C. T. VOLINSKY (1999): "Bayesian model averaging: a tutorial." *Statistical science* pp. 382–401.

JAMET, S. (2005): "Allégements généraux de cotisations sociales et emploi peu qualifié: de l'impact sectoriel à l 'effet macro-économique." *Revue française économie* **19(3)**: pp. 57–90.

JENSEN, G. A. & M. A. MORRISEY (1986): "The role of physicians in hospital production." *The Review of Economics and Statistics* pp. 432–442.

JOHNSON, G. E. (1970): "The demand for labor by educational category." *Southern Economic Journal* pp. 190–204.

KASS, R. E. & A. E. RAFTERY (1995): "Bayes factors." *Journal of the american statistical association* **90(430)**: pp. 773–795.

KATZ, L. F. & K. M. MURPHY (1992): "Changes in relative wages, 1963–1987: supply and demand factors." *The quarterly journal of economics* **107(1)**: pp. 35–78.

KAWAGUCHI, D. & Y. MORI (2016): "Why has wage inequality evolved so differently between Japan and the US? The role of the supply of college-educated workers." *Economics of Education Review* **52**: pp. 29–50.

KEARNEY, I. *et al.* (1997): "Estimating the demand for skilled labour, unskilled labour and clerical workers: a dynamic framework." *Technical report.*

Kesselman, J. R., S. H. Williamson, & E. R. Berndt (1977): "Tax credits for employment rather than investment." *The American economic review* pp. 339–349.

Kierzenkowski, R. & I. Koske (2012): "Less income inequality and more growth–are they compatible? Part 8. The drivers of labour income inequality–A literature review." .

Kim, D. I. (2005): "Growth in college education and wage differentials in Korea." *Seoul Journal of Economics* **18(2)**: p. 87.

Klenow, P. J. & A. Rodriguez-Clare (1997): "The neoclassical revival in growth economics: Has it gone too far?" *NBER macroeconomics annual* **12**: pp. 73–103.

Klotz, B., R. Madoo, & R. Hansen (1980): "A study of high and low" labor productivity" establishments in US manufacturing." In "New Developments in Productivity Measurement," pp. 239–292. University of Chicago Press.

Krusell, P., L. E. Ohanian, J.-V. Ríos-Rull, & G. L. Violante (2000): "Capital-skill complementarity and inequality: A macroeconomic analysis." *Econometrica* **68(5)**: pp. 1029–1053.

Kwack, S. Y. (2012): "Wage inequality and the contribution of capital, differential labor quality and efficiency to economic growth in Korea, 1965-2007." .

Ley, E. & M. F. Steel (2009): "On the effect of prior assumptions in Bayesian model averaging with applications to growth regression." *Journal of applied econometrics* **24(4)**: pp. 651–674.

Li, B. (2008): "The productivity of unskilled labor in multinational subsidiaries from different sources."

Li, B. (2010): "Multinational production and choice of technologies: New evidence on skill-biased technological change from China." *Economics Letters* **108(2)**: pp. 181–183.

Lindquist, M. J. (2004): "Capital–skill complementarity and inequality over the business cycle." *Review of Economic Dynamics* **7(3)**: pp. 519–540.

Madigan, D. & A. E. Raftery (1994): "Model selection and accounting for model uncertainty in graphical models using Occam's window." *Journal of the American Statistical Association* **89(428)**: pp. 1535–1546.

Malmberg, H. (2017): "Human capital and development accounting revisited." .

Manacorda, M., C. Sánchez-Páramo, & N. Schady (2010): "Changes in returns to education in Latin America: The role of demand and supply of skills." *Industrial & Labor Relations Review* **63(2)**: pp. 307–326.

Medina, C., C. M. P. Suárez *et al.* (2010): "Technical change and polarization of the labor market: Evidence for Brazil, Colombia and Mexico." *Technical report*, Banco de la República.

Mello, M. (2011): "Another look at panel estimates of the elasticity of substitution." In "33th Meeting of the Brazilian Econometric Society," .

Miller, J. N. (1993): "Tutorial review : Outliers in experimental data and their treatment." *Analyst* **118(5)**: pp. 455–461.

Mollick, A. V. (2008): "The rise of the skill premium in Mexican maquiladoras." *The Journal of Development Studies* **44(9)**: pp. 1382–1404.

Murphy, K. M., W. C. Riddell, & P. M. Romer (1998): "Wages, skills, and technology in the United States and Canada." *Technical report*, National Bureau of Economic Research.

Nissim, J. (1984): "The price responsiveness of the demand for labour by skill: British mechanical engineering: 1963-1978." *The Economic Journal* **94(376)**: pp. 812–825.

Psacharopoulos, G. & K. Hinchliffe (1972): "Further evidence on the elasticity of substitution among different types of educated labor." *Journal of Political Economy* **80(4)**: pp. 786–792.

Raftery, A. E., D. Madigan, & J. A. Hoeting (1997): "Bayesian model averaging for linear regression models." *Journal of the American Statistical Association* **92(437)**: pp. 179–191.

Razzak, W. A. & J. C. Timmins (2008): "A macroeconomic perspective on skill shortages and the skill premium in New Zealand." *Australian Economic Papers* **47(1)**: pp. 74–91.

RESHEF, A. (2007): "Heckscher-Ohlin and the global increase of skill premia: Factor intensity reversals to the rescue." *Technical report*, Mimeo: University of Virginia.

RIANO, A. (2009): "Trade, technology adoption and the rise of the skill premium in Mexico." *Technical report*, University of Nottingham Working Paper.

ROBBINS, D. J. (1996): "Evidence on trade and wages in the developing world." .

ROBINSON, J. (1969): *The economics of imperfect competition.* Springer.

ROSENTHAL, R. (1979): "The file drawer problem and tolerance for null results." *Psychological bulletin* **86(3)**: p. 638.

SANTAMARÍA, M. *et al.* (2004): *Income inequality, skills and trade: Evidence from Colombia during the 80s and 90s.* Universidad de los Andes, Facultad de Economía, CEDE.

SILVA, J. C. *et al.* (2007): "Young employment, job-skill composition and minimum wages: Evidence from a "natural experiment"." .

SMART, R. G. (1964): "The importance of negative results in psychological research." *Canadian Psychologist/Psychologie Canadienne* **5(4)**: p. 225.

SOLOW, R. M. (1956): "A contribution to the Theory of economic growth." *The quarterly journal of economics* pp. 65–94.

STANLEY, T. D. (2005): "Beyond publication bias." *Journal of Economic Surveys* **19(3)**: pp. 309–345.

SUTTON, A. J., S. DUVAL, R. TWEEDIE, K. R. ABRAMS, & D. R. JONES (2000): "Empirical assessment of effect of publication bias on meta-analyses." *Bmj* **320(7249)**: pp. 1574–1577.

TE VELDE, D. & O. MORRISSEY (2004): "Foreign direct investment, skills and wage inequality in East Asia." *Journal of the Asia Pacific Economy* **9(3)**: pp. 348–369.

THORNTON, A. & P. LEE (2000): "Publication bias in meta-analysis: Its causes and consequences." *Journal of clinical epidemiology* **53(2)**: pp. 207–216.

Tinbergen, J. (1974): "Substitution of graduate by other labour." *Kyklos* **27(2)**: pp. 217–226.

VanderPlas, J. (2014): "Frequentism and Bayesianism: A Python-driven primer." *arXiv preprint arXiv:1411.5018* .

Verdugo, G. (2014): "The great compression of the French wage structure, 1969–2008." *Labour Economics* **28**: pp. 131–144.

Wang, M. (2015): *Assigning g in Zellner's g prior for Bayesian variable selection*. Ph.D. thesis.

Wei, X., Z. Guan, G. Onel, F. Roka *et al.* (2016): "Imperfect substitution between immigrant and native farm workers in the United States." In "2016 Annual Meeting, July 31-August 2, 2016, Boston, Massachusetts," 235958. Agricultural and Applied Economics Association.

Welch, F. (1970): "Education in production." *Journal of political economy* **78(1)**: pp. 35–59.

Yang, D. (2012): "The responsiveness of firms to labor market changes in China: Evidence from micro level data." *memo, Institute of Population and Labor Economics, CASS. Central China* .

Zachłod-Jelec, M. & J. Boratyński (2016): "How large and uncertain are costs of 2030 GHG emissions reduction target for the European countries? Sensitivity analysis in a global CGE model." *Technical report*, MF Working Paper.

Zeugner, S. (2011): "Bayesian model averaging with BMS." *Tutorial to the R-package BMS 1e30* .

# Appendix A

# List of Primary Studies

Acemoglu (2002)
Angrist (1995)
Askilden & Nilsen (2005)
Autor *et al.* (2008)
Avalos & Savvides (2006)
Behar (2009)
Behar (2010)
Bergström & Panas (1992)
Berndt & Christensen (1974)
Berndt & Morrisson (1979)
Binelli (2015)
Blankenau & Cassou (2011)
Blundell *et al.* (2016)
Borghans & Ter Weel (2008)
Borjas (2003)
Borjas & Katz (2007)
Bound *et al.* (2004)
Bowles (1970)
Card & Lemieux (2001)
Card (2009)
Choi *et al.* (2005)
Ciccone & Peri (2005)
Corker & Bayoumi (1991)
Cruz *et al.* (2017)
Das (1999)

David *et al.* (2014)
Denny & Fuss (1977)
Dougherty (1972)
Dupuy & Marey (2007)
Dupuy (2007)
Fallon & Layard (1975)
Fernandez Sierra & Messina (2017)
FitzGerald *et al.* (2000)
Földvári & van Leeuwen (2006)
Freeman (1975)
Freeman & Medoff (1982)
Gallego (2012)
Gancia *et al.* (2011)
Giannarakis (2015)
Glitz & Wissmann (2017)
Goldin & Katz (2009)
Gyimah-Brempong & Gyapong (1992)
Heckman *et al.* (1998)
Jamet (2005)
Jensen & Morrisey (1986)
Johnson (1970)
Katz & Murphy (1992)
Kawaguchi & Mori (2016)
Kearney *et al.* (1997)
Kesselman *et al.* (1977)
Kim (2005)
Klenow & Rodriguez-Clare (1997)
Klotz *et al.* (1980)
Krusell *et al.* (2000)
Kwack (2012)
Li (2008)
Li (2010)
Lindquist (2004)
Madigan & Raftery (1994)
Malmberg (2017)
Manacorda *et al.* (2010)
Medina *et al.* (2010)

Mello (2011)
Mollick (2008)
Murphy *et al.* (1998)
Nissim (1984)
Psacharopoulos & Hinchliffe (1972)
Razzak & Timmins (2008)
Reshef (2007)
Riano (2009)
Robbins (1996)
Santamaría *et al.* (2004)
Silva *et al.* (2007)
Te Velde & Morrissey (2004)
Tinbergen (1974)
Verdugo (2014)
Wei *et al.* (2016)
Welch (1970)
Yang (2012)

# Appendix B

# Funnel Plot of Mean Elasticities

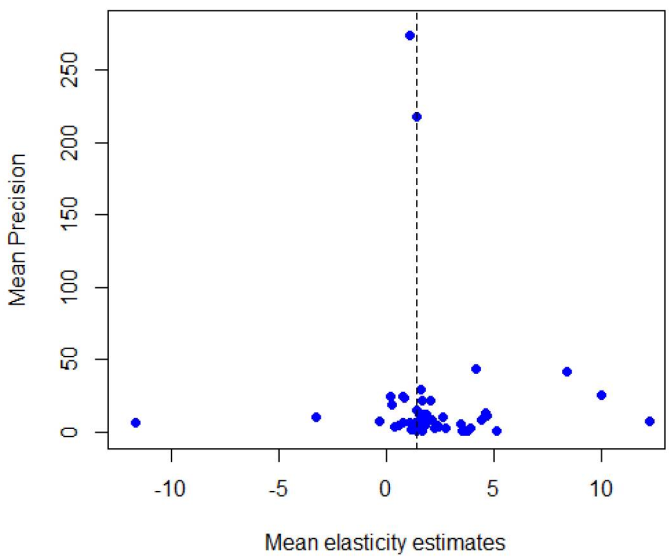



Figure B.1: Funnel Plot of Mean Estimated Elasticities

*Figure shows the funnel plot of mean elasticities from each study against mean precisions. Solid vertical line represents sample mean and dashed vertical line represents median of the elasticity estimates. 62 studies reported standard errors and 6 most extreme observations are not plotted resulting in N=56*

Table B.1: Regression-based Tests for Publication Selectivity Bias, Unweighted

*Dependent variable: estimated elasticity*

| | *All estimates* | | |
|---|---|---|---|
| | OLS Clustered | FE Clustered | ME |
| Std.error | -0.181 | 0.649 | 0.594** |
| (Bias) | (0.509) | (0.466) | (0.281) |
| Constant | 2.015*** | 1.772*** | 2.439*** |
| (Effect) | (0.531) | (0.137) | (0.280) |
| Observations | 581 | 581 | 581 |
| Studies | 62 | 62 | 62 |
| F-statistic | 0.12 | 1.94 | |
| Wald Chi-sq | | | 4.46 |

| | *Unpublished estimates* | | | *Published estimates* | | |
|---|---|---|---|---|---|---|
| | OLS Clustered | FE Clustered | ME | OLS Clustered | FE Clustered | ME |
| Std.error | 0.065 | 1.163*** | 1.112*** | -0.231 | -1.654 | -0.881 |
| (Bias) | (0.533) | (0.649) | (0.269) | (0.968) | (1.475) | (0.637) |
| Constant | 1.367*** | 1.024*** | 1.921*** | 2.718*** | 3.101*** | 3.13*** |
| (Effect) | (5.607) | (7.863) | (10.849) | (8.999) | (0.420) | (13.951) |
| Observations | 318 | 318 | 318 | 263 | 263 | 263 |
| Studies | 26 | 26 | 26 | 36 | 36 | 36 |
| F-statistic | 0.01 | 321.01 | | 0.06 | 1.26 | |
| Wald Chi-sq | | | 17.1 | | | 1.91 |

*Note: Estimated elasticities and their standard errors are winsorized.*

# Appendix C

# Correlation matrix

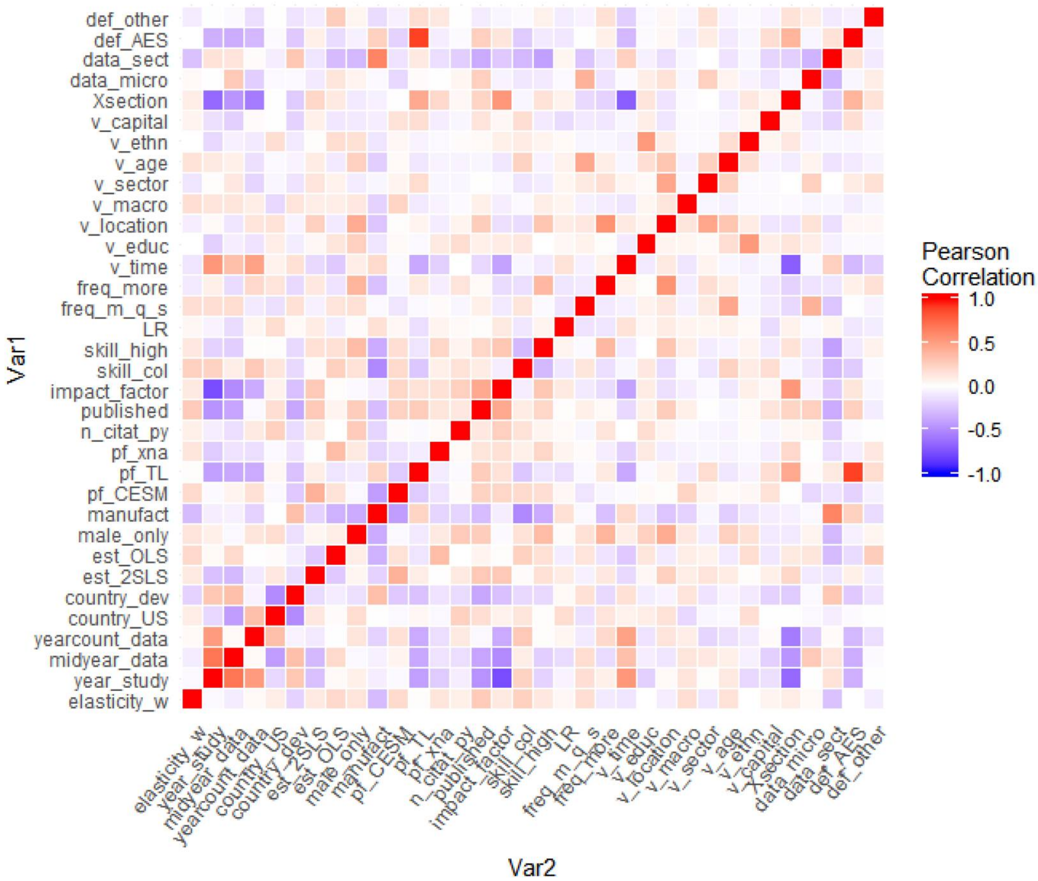



Figure C.1: Correlation matrix for explanatory variables in heterogeneity analysis

| *Variable 1* | *Variable 2* | *Correlation* |
|---|---|---|
| def_AES | pf_TL | 0.89 |
| manufacturing | data_sec | 0.62 |
| midyear_data | year_study | 0.69 |
| Xsection | year_study | -0.65 |
| Xsection | v_time | -0.70 |
| impact_factor | year_study | -0.77 |

Table C.1: Explanatory variables with the absolute value of correlation $> 0.6$

# Appendix D

# BMA Diagnostics

Diagnostics for Bayesian model averaging assuming normal prior distribution.

| | | | |
|---|---|---|---|
| *Mean no. of regressors* | *Draws* | *Burn-ins* | *Time* |
| 16.98 | $2x10^6$ | $10^6$ | 4.34 mins |
| *No. models visited* | *Modelspace* | *Visited* | *Topmodels* |
| 493722 | $4.3x10^9$ | 0.011% | 85% |
| *Corr PMP* | *No. Obs* | *Model prior* | *g-Prior* |
| 0.9995 | 684 | random | BRIC |
| *Shrinkage-Stats* | | | |
| av.=0.999 | | | |

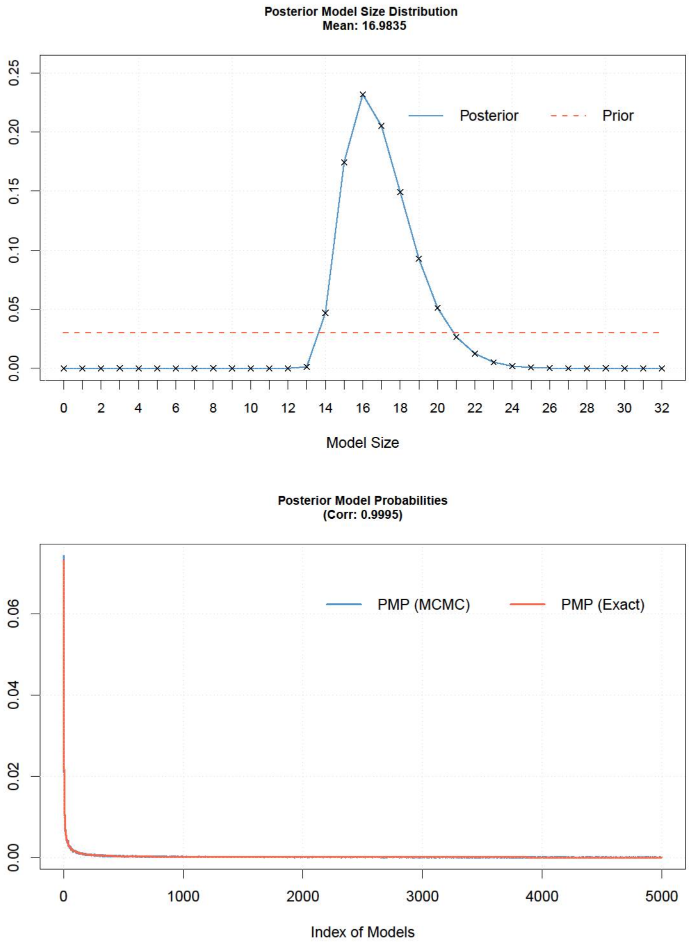



Figure D.1: BMA, Prior and Posterior Model Probabilities

Figure D.2: Posterior Densities for Variables with PIP> 0.5

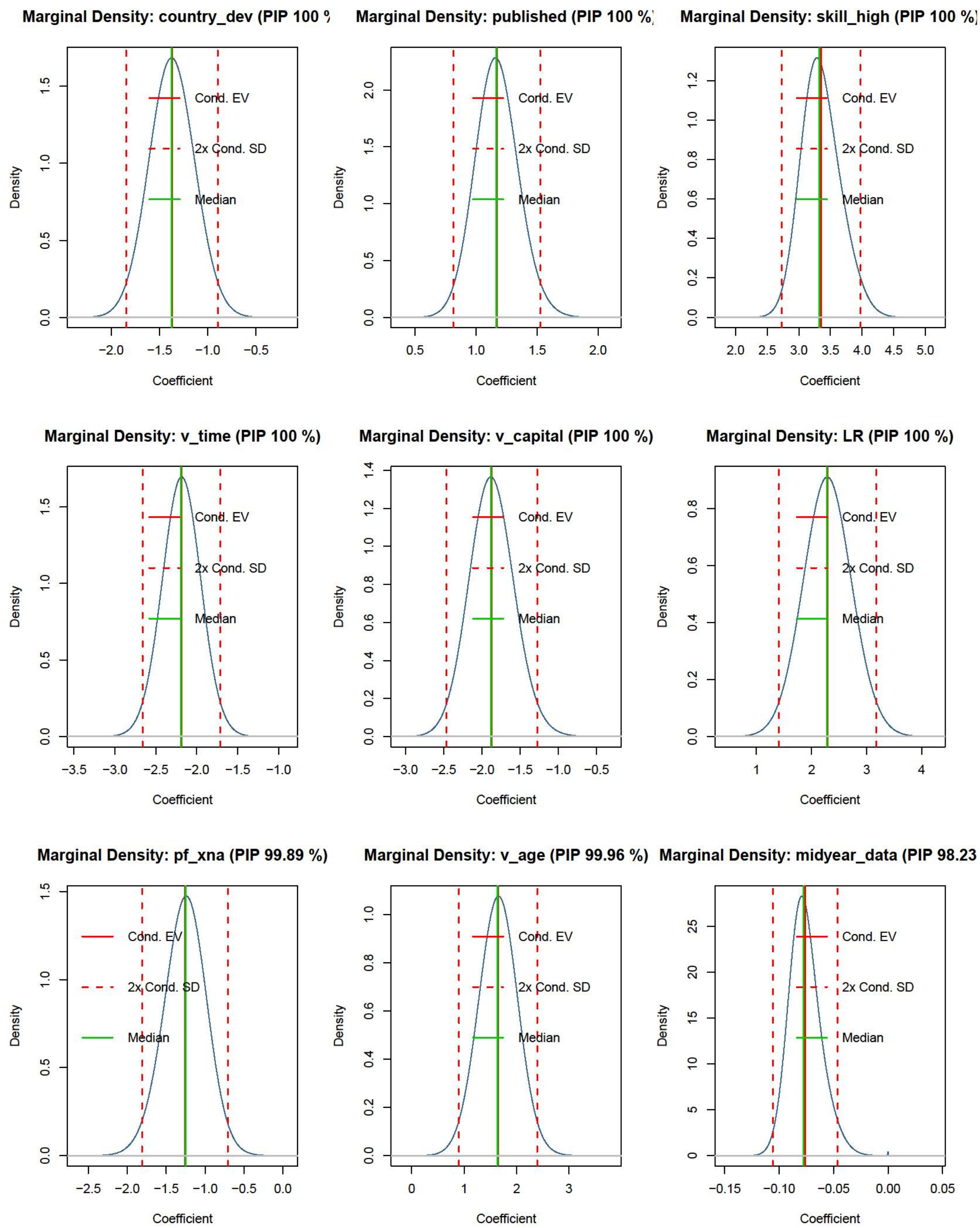
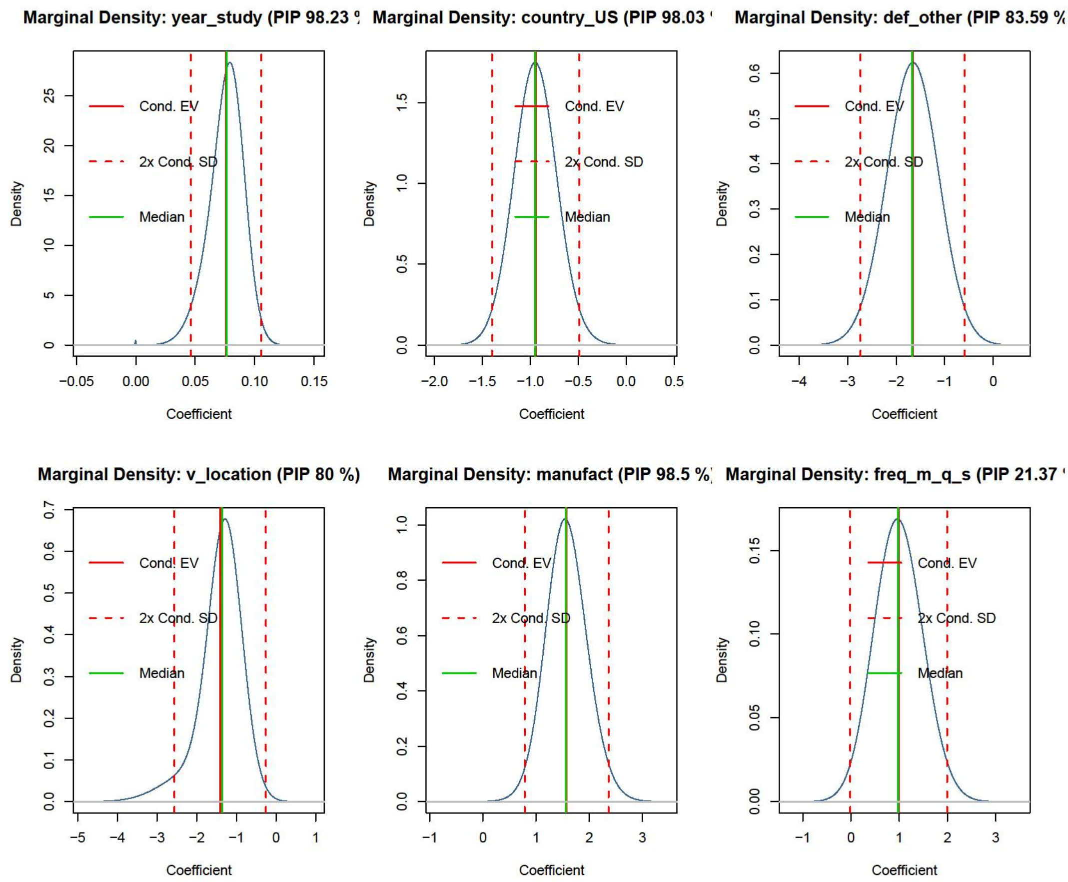


Figure D.3: Continued: Posterior Densities for Variables with PIP> 0.5



# Appendix E

# BMA Robustness Check

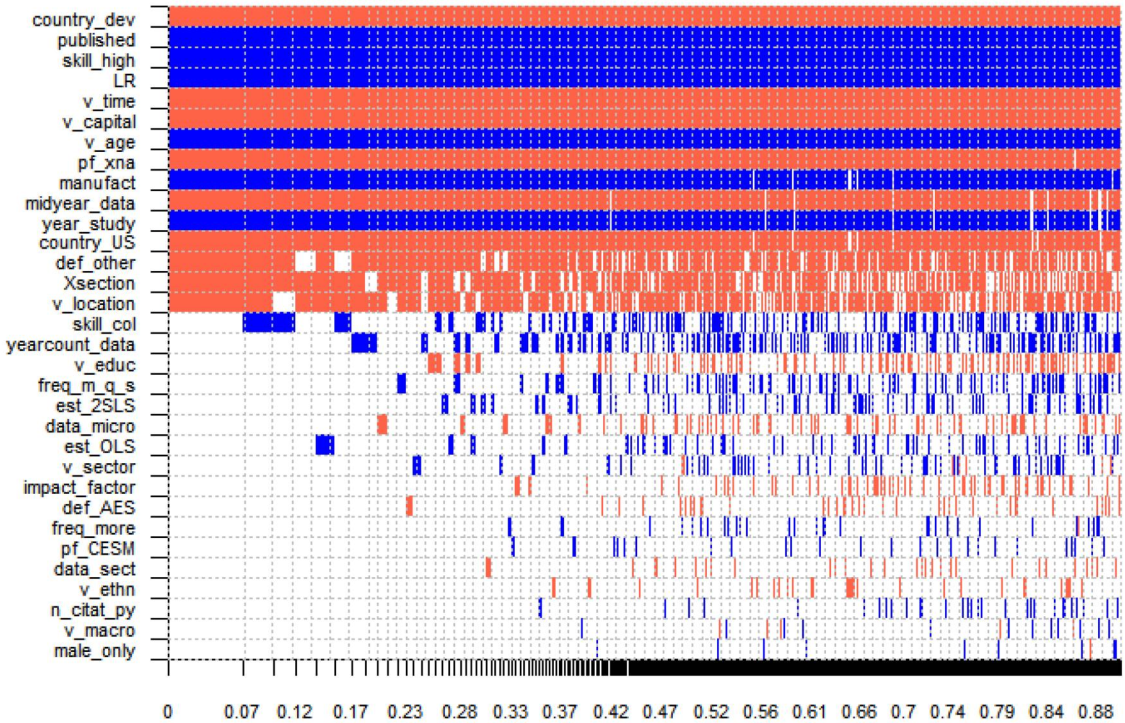



Figure E.1: BMA, *mprior=uniform, g=BRIC*

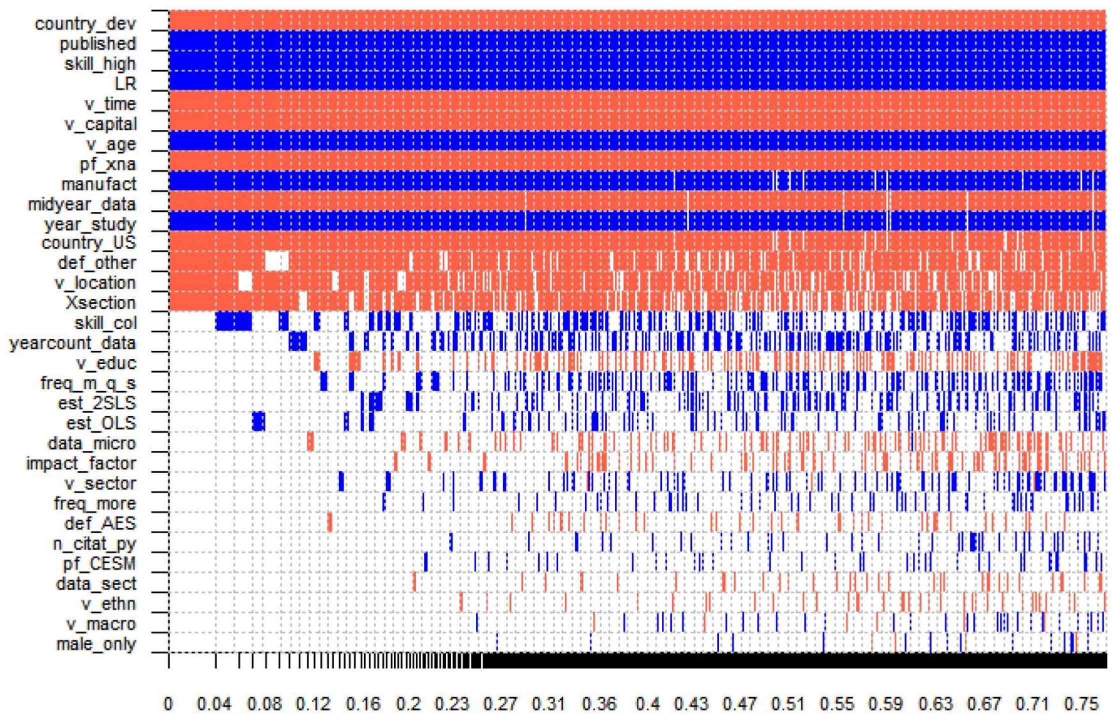



Figure E.2: BMA, *mprior=random, g=UIP*

Table E.1: Heterogeneity-Alternative BMAs

| | *BMA: mprior=uniform* | | | *BMA: g=UIP* | | |
|---|---|---|---|---|---|---|
| | Post Mean | Post St.D. | PIP | Post Mean | Post St.D. | PIP |
| *Real Factors* | | | | | | |
| country_dev | -1.371 | 0.237 | 1.000 | -1.370 | 0.237 | 1.000 |
| country_US | -0.933 | 0.258 | 0.981 | -0.910 | 0.266 | 0.978 |
| midyear_data | -0.075 | 0.018 | 0.981 | -0.073 | 0.018 | 0.986 |
| manufacturing | 1.555 | 0.436 | 0.984 | 1.549 | 0.443 | 0.986 |
| male_only | 0.001 | 0.035 | 0.024 | 0.002 | 0.040 | 0.028 |
| *Elasticity type* | | | | | | |
| def_AES | -0.045 | 0.205 | 0.068 | -0.059 | 0.236 | 0.088 |
| def_other | -1.371 | 0.795 | 0.826 | -1.454 | 0.749 | 0.873 |
| longrun | 2.275 | 0.438 | 1.000 | 2.330 | 0.448 | 1.000 |
| pf_xna | -1.245 | 0.276 | 0.999 | -1.282 | 0.283 | 0.999 |
| pf_CESM | 0.015 | 0.079 | 0.057 | 0.067 | 0.085 | 0.067 |

Table E.2: Continued: Heterogeneity-Alternative BMAs

| | Post Mean | Post St.D. | PIP | Post Mean | Post St.D. | PIP |
|---|---|---|---|---|---|---|
| *Data* | | | | | | |
| data_micro | -0.072 | 0.200 | 0.151 | -0.080 | 0.212 | 0.168 |
| data_sect | -0.021 | 0.112 | 0.054 | -0.023 | 0.118 | 0.063 |
| Xsection | -0.770 | 0.458 | 0.806 | -0.745 | 0.461 | 0.797 |
| yearcount_data | 0.007 | 0.012 | 0.298 | 0.007 | 0.013 | 0.344 |
| freq_m_q_s | 0.179 | 0.434 | 0.185 | 0.282 | 0.523 | 0.280 |
| freq_more | 0.025 | 0.129 | 0.060 | 0.091 | 0.174 | 0.091 |
| skill_high | 3.345 | 0.309 | 1.000 | 3.373 | 0.318 | 1.000 |
| skill_col | 0.167 | 0.270 | 0.327 | 0.201 | 0.286 | 0.392 |
| *Estimation* | | | | | | |
| est_2SLS | 0.102 | 0.280 | 0.153 | 0.149 | 0.324 | 0.222 |
| est_OLS | 0.050 | 0.155 | 0.125 | 0.072 | 0.186 | 0.171 |
| v_time | -2.178 | 0.235 | 1.000 | -2.208 | 0.239 | 1.000 |
| v_capital | -1.872 | 0.296 | 1.000 | -1.864 | 0.302 | 1.000 |
| v_age | 1.644 | 0.371 | 0.999 | 1.655 | 0.384 | 0.999 |
| v_location | -1.088 | 0.752 | 0.785 | -1.213 | 0.776 | 0.839 |
| v_educ | -0.107 | 0.259 | 0.185 | -0.180 | 0.327 | 0.288 |
| v_sector | 0.131 | 0.528 | 0.105 | 0.234 | 0.667 | 0.162 |
| v_ethn | -0.036 | 0.211 | 0.050 | -0.036 | 0.208 | 0.055 |
| v_macro | 0.008 | 0.090 | 0.036 | 0.019 | 0.127 | 0.052 |
| *Publication* | | | | | | |
| published | 1.167 | 0.176 | 1.000 | 1.177 | 0.183 | 1.000 |
| year_study | 0.075 | 0.018 | 0.981 | 0.072 | 0.018 | 0.985 |
| impact_factor | -0.013 | 0.050 | 0.098 | -0.026 | 0.069 | 0.164 |
| n_citat_py | 0.000 | 0.001 | 0.047 | 0.000 | 0.001 | 0.083 |
| (Constant) | -0.017 | | 1.000 | -0.016 | | 1.000 |

*Notes: N=684. Weighted by inverse of the number of estimates per study. Elasticity is winsorized.*